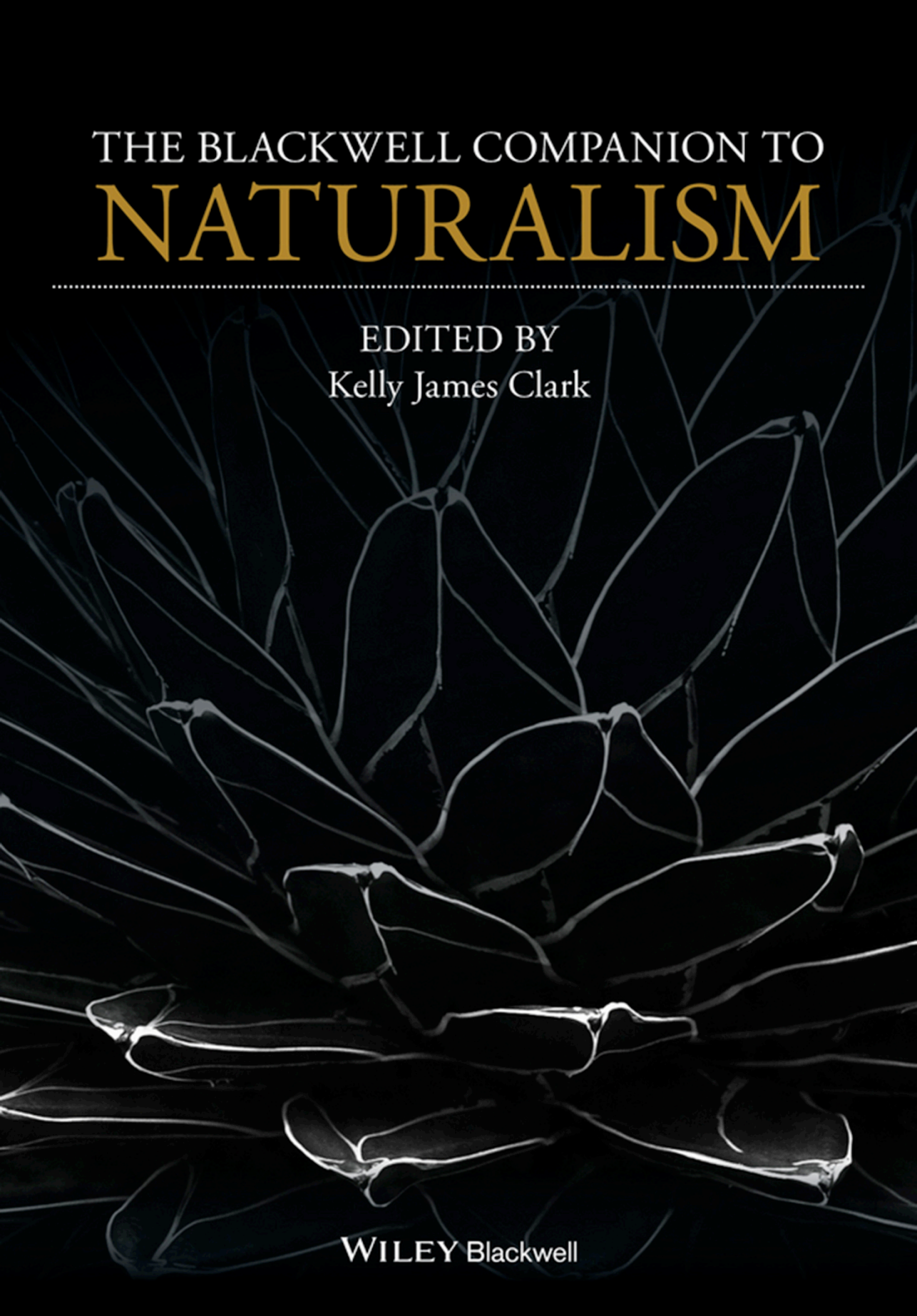
THE BLACKWELL COMPANION TO NATURALISM

EDITED BY
Kelly James Clark



WILEY Blackwell

The Blackwell Companion to Naturalism

Blackwell Companions to Philosophy

This outstanding student reference series offers a comprehensive and authoritative survey of philosophy as a whole. Written by today's leading philosophers, each volume provides lucid and engaging coverage of the key figures, terms, topics, and problems of the field. Taken together, the volumes provide the ideal basis for course use, representing an unparalleled work of reference for students and specialists alike.

Already published in the series:

1. The Blackwell Companion to Philosophy, Second Edition
Edited by Nicholas Bunnin and Eric Tsui-James

2. A Companion to Ethics
Edited by Peter Singer

3. A Companion to Aesthetics, Second Edition
Edited by Stephen Davies, Kathleen Marie Higgins, Robert Hopkins, Robert Stecker, and David E. Cooper

4. A Companion to Epistemology, Second Edition
Edited by Jonathan Dancy, Ernest Sosa and Matthias Steup

5. A Companion to Contemporary Political Philosophy (two-volume set), Second Edition
Edited by Robert E. Goodin and Philip Pettit

6. A Companion to Philosophy of Mind
Edited by Samuel Guttenplan

7. A Companion to Metaphysics, Second Edition
Edited by Jaegwon Kim, Ernest Sosa and Gary S. Rosenkrantz

8. A Companion to Philosophy of Law and Legal Theory, Second Edition
Edited by Dennis Patterson

9. A Companion to Philosophy of Religion, Second Edition
Edited by Charles Taliaferro, Paul Draper, and Philip L. Quinn

10. A Companion to the Philosophy of Language
Edited by Bob Hale and Crispin Wright

11. A Companion to World Philosophies
Edited by Eliot Deutsch and Ron Bontekoe

12. A Companion to Continental Philosophy
Edited by Simon Critchley and William Schroeder

13. A Companion to Feminist Philosophy
Edited by Alison M. Jaggar and Iris Marion Young

14. A Companion to Cognitive Science
Edited by William Bechtel and George Graham

15. A Companion to Bioethics, Second Edition
Edited by Helga Kuhse and Peter Singer

16. A Companion to the Philosophers
Edited by Robert L. Arrington

17. A Companion to Business Ethics
Edited by Robert E. Frederick

18. A Companion to the Philosophy of Science
Edited by W. H. Newton-Smith

19. A Companion to Environmental Philosophy
Edited by Dale Jamieson

20. A Companion to Analytic Philosophy
Edited by A. P. Martinich and David Sosa

21. A Companion to Genethics
Edited by Justine Burley and John Harris

22. A Companion to Philosophical Logic
Edited by Dale Jacquette

23. A Companion to Early Modern Philosophy
Edited by Steven Nadler

24. A Companion to Philosophy in the Middle Ages
Edited by Jorge J. E. Gracia and Timothy B. Noone

25. A Companion to African-American Philosophy
Edited by Tommy L. Lott and John P. Pittman

26. A Companion to Applied Ethics
Edited by R. G. Frey and Christopher Heath Wellman

27. A Companion to the Philosophy of Education
Edited by Randall Curren

28. A Companion to African Philosophy
Edited by Kwasi Wiredu

29. A Companion to Heidegger
Edited by Hubert L. Dreyfus and Mark A. Wrathall

30. A Companion to Rationalism
Edited by Alan Nelson

31. A Companion to Pragmatism
Edited by John R. Shook and Joseph Margolis

32. A Companion to Ancient Philosophy
Edited by Mary Louise Gill and Pierre Pellegrin

33. A Companion to Nietzsche
Edited by Keith Ansell Pearson

34. A Companion to Socrates
Edited by Sara Ahbel-Rappe and Rachana Kamtekar

35. A Companion to Phenomenology and Existentialism
Edited by Hubert L. Dreyfus and Mark A. Wrathall

36. A Companion to Kant
Edited by Graham Bird

37. A Companion to Plato
Edited by Hugh H. Benson

38. A Companion to Descartes
Edited by Janet Broughton and John Carriero

39. A Companion to the Philosophy of Biology
Edited by Sahotra Sarkar and Anya Plutynski

40. A Companion to Hume
Edited by Elizabeth S. Radcliffe

41. A Companion to the Philosophy of History and Historiography
Edited by Aviezer Tucker

42. A Companion to Aristotle
Edited by Georgios Anagnostopoulos

43. A Companion to the Philosophy of Technology
Edited by Jan-Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks

44. A Companion to Latin American Philosophy
Edited by Susana Nuccetelli, Ofelia Schutte, and Otávio Bueno

45. A Companion to the Philosophy of Literature
Edited by Garry L. Hagberg and Walter Jost

46. A Companion to the Philosophy of Action
Edited by Timothy O'Connor and Constantine Sandis

47. A Companion to Relativism
Edited by Steven D. Hales

48. A Companion to Hegel
Edited by Stephen Houlgate and Michael Baur

49. A Companion to Schopenhauer
Edited by Bart Vandenabeele

50. A Companion to Buddhist Philosophy
Edited by Steven M. Emmanuel

51. A Companion to Foucault
Edited by Christopher Falzon, Timothy O'Leary, and Jana Sawicki

52. A Companion to the Philosophy of Time
Edited by Heather Dyke and Adrian Bardon

53. A Companion to Donald Davidson
Edited by Ernest Lepore and Kirk Ludwig

54. A Companion to Rawls
Edited by Jon Mandle and David Reidy

55. A Companion to W.V.O Quine
Edited by Gilbert Harman and Ernest Lepore

56. A Companion to Derrida
Edited by Zeynep Direk and Leonard Lawlor

57. A Companion to David Lewis
Edited by Barry Loewer and Jonathan Schaffer

58. A Companion to Kierkegaard
Edited by Jon Stewart

59. A Companion to Locke
Edited by Matthew Stuart

60. The Blackwell Companion to Hermeneutics
Edited by Niall Keane and Chris Lawn

61. A Companion to Ayn Rand
Edited by Allan Gotthelf and Gregory Salmieri

62. The Blackwell Companion to Naturalism
Edited by Kelly James Clark

THE BLACKWELL COMPANION TO NATURALISM

Edited by

Kelly James Clark

WILEY Blackwell

This edition first published 2016
© 2016 John Wiley & Sons, Inc

Registered Office
John Wiley & Sons, Inc, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of Kelly James Clark to be identified as the author of the editorial material in this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Names: Clark, Kelly James, 1956– editor.
Title: A companion to naturalism / edited by Kelly James Clark.
Description: 1 [edition]. | Hoboken : John Wiley & Sons, Inc., 2016. | Includes index.
Identifiers: LCCN 2015041263 | ISBN 9781118657607 (hardback)
Subjects: LCSH: Naturalism. | BISAC: PHILOSOPHY / Metaphysics.
Classification: LCC B828.2.C58 2016 | DDC 146–dc23 LC record available at http://lccn.loc.gov/2015041263

A catalogue record for this book is available from the British Library.

Set in 10/12.5 Minion by SPi Global, Pondicherry, India

1 2016

To Michael Murray, friend and philosopher

Contents

	List of Contributors	ix
	Preface	xii
1	Naturalism and its Discontents Kelly James Clark	1
2	Naturalizing Ethics Owen Flanagan, Hagop Sarkissian, and David Wong	16
3	Naturalism in the Continental Tradition Keith Ansell Pearson and John Protevi	34
4	The Naturalism Question in Feminism Ásta Sveinsdóttir	49
5	On Naturalistic Metaphysics Thomas M. Crisp	61
6	Naturalism and Realism in the Philosophy Science Matteo Morganti	75
7	Naturalism without Scientism P. Kyle Stanford	91
8	"The Horrid Doubt": Naturalism and Evolutionary Biology Valerie Gray Hardcastle	109
9	Naturalism and Antinaturalism in the Sociology of Science Dorothea Olkowski	124
10	Why Methodological Naturalism? Hans Halvorson	136
11	Naturalism and the Question of Realism Drew Khlentzos	150
12	Non-Naturalistic Metaphysics Hud Hudson	168
13	Naturalism and Physicalism Barbara Gail Montero and David Papineau	182
14	Natural Mind Brian L. Keeley	196

15 Naturalism and Dualism 209
Joseph Levine
16 Epistemological Naturalisms 220
C.S.I. Jenkins
17 Dewey, Naturalism, and the Problem of Knowledge 234
Douglas McDermid
18 Truth and Naturalism 246
Douglas Edwards, Filippo Ferrari, and Michael P. Lynch
19 Against Naturalism about Truth 262
Berit Brogaard
20 Mathematics and Metaphysical Naturalism 277
Gideon Rosen
21 Naturalism and Mathematics: Some Problems 289
Jeffrey W. Roland
22 Naturalism and Free Will 305
Neil Levy
23 Free Will and Naturalism: How to Be a Libertarian, and a Naturalist Too 319
Kevin Timpe and Jonathan D. Jacobs
24 Does the New Wave in Moral Psychology Sink Kant? 336
Valerie Tiberius
25 Naturalism in Metaethics 351
Jussi Suikkanen
26 Evolution and Moral Naturalism 369
Richard Joyce
27 Scientific Naturalism and the Explanation of Moral Beliefs:
Challenging Evolutionary Debunking 386
William J. FitzPatrick
28 What's to be Said for Moral Non-Naturalism? 401
Terence D. Cuneo
29 Naturalism and Moral Psychology 416
Christian B. Miller
30 Militant Modern Atheism 435
Philip Kitcher
31 Why Naturalism Cannot Account for Natural Human Rights 447
Nicholas Wolterstorff
32 Cognitive and Evolutionary Approaches to Religion 462
Robert N. McCauley
33 The Naturalness of Religious Belief: Epistemological Implications 481
Helen De Cruz
34 Naturalism in Indian Philosophy 494
Amita Chatterjee
35 The Natural History of Shame and its Modification by Confucian Culture 512
Ryan Nichols

Index 528

List of Contributors

Ásta Sveinsdóttir
San Francisco State University
San Francisco, CA, USA

Berit Brogaard
University of Miami, Miami, FL, USA
University of Oslo
Oslo, Norway

Amita Chatterjee
School of Cognitive Science
Jadavpur University
Kolkata, India

Kelly James Clark
Kaufman Institute
Grand Valley State University
Allendale, MI, USA

Thomas M. Crisp
Biola University
Brea, CA, USA

Terence D. Cuneo
University of Vermont
Burlington, VT, USA

Helen De Cruz
Faculty of Humanities
VU University Amsterdam
Amsterdam, The Netherlands

Douglas Edwards
Hamilton College
Clinton, NY, USA

Filippo Ferrari
University of Aberdeen
Aberdeen, UK

William J. FitzPatrick
University of Rochester
Rochester, NY, USA

Owen Flanagan
Duke University
Durham, NC, USA

Hans Halvorson
Princeton University
Princeton, NJ, USA

Valerie Gray Hardcastle
University of Cincinnati
California, KY, USA

Hud Hudson
Western Washington University
Bellingham, WA, USA

Jonathan D. Jacobs
St. Louis University
Saint Louis, MO, USA

C.S.I. Jenkins
University of British Columbia
Vancouver, BC, Canada

Richard Joyce
Victorian University of Wellington
Wellington, New Zealand

Brian L. Keeley
Pitzer College
Claremont, CA, USA

Drew Khlentzos
Australian Research Centre of Excellence in Cognition and its Disorders and Department of Linguistics
Macquarie University
Australia

Philip Kitcher
Columbia University
New York, NY, USA

Joseph Levine
UMass Amherst
Leverett, MA, USA

Neil Levy
Oxford University
St. Kilda East, Australia

Michael P. Lynch
University of Connecticut
Mansfield Center, CT, USA

Robert N. McCauley
Emory University
Atlanta, GA, USA

Douglas McDermid
Trent University
Peterborough, ON, Canada

Christian B. Miller
Wake Forest University
Winston-Salem, NC, USA

Barbara Gail Montero
CUNY
New York, NY, USA

Matteo Morganti
Department of Philosophy
Communication and Performing Arts
University of Rome 'Roma TRE'
Rome, Italy

Ryan Nichols
Department of Philosophy
California State University, Fullerton
Fullerton, CA, USA

Dorothea Olkowski
University of Colorado, Colorado Spring
Colorado Springs, CO, USA

David Papineau
CUNY, New York, NY, USA
Kings College, London
London, UK

Keith Ansell Pearson
University of Warwick
Coventry, UK

John Protevi
Louisiana State University
Baton Rouge, LA, USA

Jeffrey W. Roland
Louisiana State University
Baton Rouge, LA, USA

Gideon Rosen
Princeton University
Princeton, NJ, USA

Hagop Sarkissian
Department of Philosophy,
Baruch College
CUNY
New York, NY, USA

P. Kyle Stanford
University of California, Department of Logic and Philosophy of Science
Irvine, CA, USA

Jussi Suikkanen
University of Birmingham
Edgbaston
Birmingham, UK

Valerie Tiberius
University of Minnesota
Minneapolis, MN, USA

Kevin Timpe
Northwest Nazarene College
Nampa, ID, USA

Nicholas Wolterstorff
Yale University
Grand Rapids, MI, USA

David Wong
Duke University
Durham, NC, USA

Preface

A few years ago, I read Owen Flanagan, Hagop Sarkissian and David Wong's essay, "Naturalizing Ethics."[1] Unlike much recent philosophy, it was expansive, bold, and incomplete. I mean all of those as compliments. It laid out, very clearly, their understanding of naturalism, and then explored the implications of naturalism thus understood for ethics. As they are keenly aware, defending naturalism and a naturalized ethics requires considerably more attention than a single essay affords. But sometimes philosophers should think broadly, worldviewishly, and suggestively; leave the pesky details to be filled in later (perhaps by others).

I wrote to Owen and asked if he'd be willing to let me use that essay as the opening salvo in a contrasting-views book on naturalism and, after consulting with Hagop and David, he generously agreed. I approached Blackwell with the completed project and they encouraged me to think bigger. By the time we were done thinking, the idea for the *Blackwell Companion to Naturalism* had emerged. It would define and develop naturalism. It would also offer criticisms, friendly and otherwise. And it would contain essays, like the original piece, that would speak to a wide audience in broad and suggestive ways. A few years later, voila, the book is now complete.

"Voila" makes it sound easier than it was. I felt at times, in the lyrics of the Bob Seger tune, that I was running against the wind; that sounds overly dramatic, but I was faced with difficult decisions about "what to leave in, what to leave out." Naturalism and ethics seemed one of the contemporary pressure points for assessing the prospects and problems of naturalism; there are more essays on this topic than any other. More topics emerged, such as truth, knowledge, science, metaphysics, mind, social–political philosophy, and religion. I sought defenders on the extremes of these views – not to be provocative, but to help the reader get a better sense of a lively, ongoing, and even important debate. Of course, on the running-against-the-wind score, there were scofflaws that missed deadlines and ignored my entreaties!

The completed volume offers a snapshot of a moving target. How should we conceive of naturalism, and what are its consequences for, say, ethics or knowledge? It's a moving target partly because naturalism constitutes more than a denial of supernaturalism; it also includes

1 Reprinted as Chapter 2 in this volume.

some sort of special allegiance to science. But science has not, as far as we know, achieved its final form, and there is much that is as yet scientifically unknown. So we simply don't know exactly what allegiance to science entails for this or that philosophical topic. Naturalism and its consequences is a work in progress.

I am grateful for the kind support and encouragement of my wife, Susan. This work was generously supported by a grant from the John Templeton Foundation.

Kelly James Clark

1

Naturalism and its Discontents

KELLY JAMES CLARK

> Every philosopher has taken his stand on a sort of dumb conviction that the truth must lie in one direction rather than another.
>
> William James

Introduction

In 1922, the philosopher Roy Wood Sellars proclaimed, "We are all naturalists now" (1922, vii). If by "we" Sellars meant fellow philosophers or academics, his triumphant declaration, though untrue then, is closer to being true now. At the turn of the 21st century, naturalism is the reigning orthodox assumption of most faculty in most universities across the English-speaking world. In the discipline of philosophy, philosophers increasingly identify themselves as naturalists.

What exactly is naturalism? Naturalism, we will learn, admits of no single, simple definition and comes in a wide variety of shapes and sizes (depending, for example, on its commitment to the natural sciences). After distinguishing ontological or metaphysical naturalism from the considerably more modest methodological naturalism, I will discuss the historical development of ontological naturalism, as well as arguments for or against naturalism generally. Before concluding, I will take moral goodness and badness as a case study of the problems and prospects for ontological naturalism.

What Is Naturalism?

It is impossible to offer a single precise definition of "naturalism," one that captures everything that goes by the name. Defined too narrowly, it leaves out wide swaths of human thought and experience; defined too broadly, it includes many things that naturalists hope to exclude. A better approach, then, is to consider various understandings of naturalism, as

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

well as naturalism's historical development and recent rise. As we will see, naturalism has been understood in a great many ways. Given the diversity of understandings, we will cast our nets widely and consider the many different views and ideas that fall under the banner of "naturalism," while noting uniting features that (at least most) naturalists share.

Naturalists, as might be expected, give pride of place to the natural world, to nature, and are dubious or even dismissive of claims that go beyond the natural world: so-called "supernatural" claims. These include most importantly the belief that God exists, but also beliefs in other sorts of non-natural or spiritual entities, such as ghosts and, for most naturalists, the human soul or spirit (as an entity that is independent from the brain), as well as non-natural or spiritual powers such as *qi* and astrological forces. So naturalists are typically atheists who reject, among other things, mind–body dualism. As such, naturalism is an *ontology*: it makes claims about what exists (and, perhaps more importantly, what does not). This view is frequently called *metaphysical naturalism*.

This rough formulation doesn't tell us too much about the commitments of ontological or metaphysical naturalists. Nevertheless, there is one way in which it is helpful. That is, we can think of naturalism as standing in opposition to allegedly "spooky" – spiritual or theistic – understandings of the world and reality. Naturalists, then, reject any appeals to divine or supernatural entities or powers in their explanation of the world. Examples include, as we've seen, the claim that God created the universe and everything it contains and that human persons are a composition of soul and body. Naturalism came to prominence as an alternative to theism as a way of explaining the world.

As with all philosophies and philosophers, there is disagreement among naturalists. What precisely is "the natural world"? Some claim that it consists ultimately of subatomic particles (Rosenberg 2011), the ultimate and indivisible material reality. If the world were constructed out of material atoms, then naturalism would be synonymous with materialism (the view that everything is matter). But science has gone beyond (or "beneath") atoms in its understanding of nature. Perhaps nature is made up not of atoms, but instead of matter and energy, or matter/energy in its various manifestations and relations. Or perhaps matter is reducible to waves or packets of energy, or even some sort of mental stuff. The point is, we don't really know just what constitutes the natural world (and thus we don't really know what lies beyond the limits of the natural world).

To be sure, naturalism is not by definition materialistic. Logical positivism, which arose out of philosophical naturalism, rejected attempts to go beyond the domain of human experience. Positivism's severe empiricism restricted knowledge (and knowledge-based reality) to sensory experience (usually called "sense data") and logical constructions of sense experience. Physical objects were considered no more than bundles of sense data. A chair, for example, is not a solid, enduring physical object (out there, in the mind-independent world); a chair, instead, is a collection of actual and possible experiences or sensations. Since experience is mental, reality for the positivist is mental. In short, for the logical positivist, naturalism (when allied with strict empiricism) entails idealism (the idea that reality is fundamentally mental).[1]

It's also hard to know what counts as supernatural. Most naturalists would say that God, spirits, angels, demons, and souls are supernatural entities. But, beyond these, there is little consensus. What about such apparently non-natural, immaterial things as numbers, moral facts, merely possible beings (such as unicorns or fictional characters), and other abstract

1 This inference would not follow if sense data were physical objects.

entities (like the concepts "red" and "true")? Where in the natural world do we find "Murder is wrong," Pegasus, or "2 + 2 = 4" (or the number 2)?

On these matters, as well as many others, naturalists themselves disagree. Even some atheist philosophers resist the temptation to naturalism. They don't believe in God or gods, all right, but they do believe in such non-natural, immaterial entities or facts as numbers and morality, so they don't believe they can in good conscience call themselves naturalists. Some naturalists, on the other hand, believe that immaterial entities such as numbers and other abstract entities are part and parcel of the natural world (or will eventually be shown to be so).

For the discussion that follows, we will (with provisos along the way) understand naturalism in the metaphysical or ontological sense that *everything that exists is included in the natural world; there are no supernatural entities*. We will from now on call metaphysical or ontological naturalism, unless otherwise noted, "Naturalism."

Naturalism and Science

Some contemporary Naturalists go further, holding that scientific inquiry is *the* way of knowing in general and that the finally established results of science (especially physics) determine our view of reality. According to these Naturalists, *the sciences are the only guide to understanding reality*. Wilfrid Sellars, in a playful paraphrase of Protagoras, writes: "Science is the measure of all things, of what is that it is, and of what is not, that it is not" (Sellars 1963, 173). Such Naturalists have a correspondingly low regard for nonscientific forms of inquiry, such as those that appeal to intuition or tradition, or to religious experience and texts. Because such Naturalists believe science is the sole guide to understanding reality, they think we should apply the methodology of science to all domains of inquiry. Recent debates in metaphysics divide over those who think metaphysics should be studied/advanced independent of any considerations of science and those more naturalistically inclined who think metaphysics should take its cues from science.[2]

Scientific Naturalism constitutes a role reversal for traditional approaches to philosophy, which have relied on *a priori* theorizing about the nature and extent of reality. Philosophy comes first in the order of inquiry, on this traditional view, and science comes second; accordingly, philosophy has authority over science. But philosophy is no longer considered the queen of the sciences, sitting in judgment over scientific claims; philosophy is often now believed to be the servant of the sciences, taking its dictates as the sober and ultimate truth about the nature of Reality. Science, on this view, has authority over philosophy. Philosophy may help us understand the foundations and methods of the sciences, which scientists employ in their various practices, but science sets the limits and nature of human inquiry. Philosophy, on this view, simply accedes to the dictates of science.

Privileging science has much to commend it: there is no other domain of human inquiry that has been so remarkably successful in understanding the world and achieving rational consensus. Tradition, authority, and Holy Writ, for example, have failed to produce the rational consensus that we find in science (or any rational consensus whatsoever). Even those who affirm this or that text as Scripture find themselves with substantial disagreement

2 See Chalmers, Manley, and Wasserman (2009) for the former view and Ross, Ladyman, and Kincaid (2013) for the latter.

as to what their Scripture teaches. Science or scientific inquiry offers what religion promised but has failed to offer: a method of inquiry for attaining rational consensus.

More than consensus, though, science also seems to be uniquely capable of attaining the truth: the universal law of gravitation, for example, or our sun-centered planetary system, or the age of the universe. More controversially, but I think undeniably, it has shown that all species, including the human species, have evolved from a single ancestor.

Finally, commitment to scientific modes of inquiry brings to Naturalists a certain open-mindedness: one should follow science wherever it leads, even if it challenges or contradicts some of one's fundamental and cherished assumptions about the nature of reality. A scientific Naturalism may claim that nature is all that exists but also hold that nature is whatever will be disclosed by the ideal natural sciences. Since contemporary science is not, at least as far as we can tell, the ideal science, at this point in human history we may know very little about nature. So the scientific Naturalist remains open to understanding nature as science continues to disclose it.

Despite these marks in favor of naturalism, there are reasons to think that science is not able to adjudicate truth all by itself. For example, science must assume the mathematical truths that are so essential to its success. Where in the natural world can one find numbers? And while science has shown the dependence of the mental on the physical, it has not yet shown, even in principle, that consciousness is reducible to the brain or to brain states. Finally, and even more pressing existentially, is that science, the world of facts, seems devoid of everything that we so deeply value: goodness, meaning, and purpose. The physicist Erwin Schroedinger is a critic of science-based Naturalism: "I am very astonished that the scientific picture of the real world around me is very deficient. It gives a lot of factual information, puts all our experiences in a magnificently consistent order, but is ghastly silent about all and sundry that is really near to our heart, that really matters to us. It cannot tell us a word about red and blue, bitter and sweet, physical pain and physical delight; it knows nothing of beautiful and ugly, good or bad, God and eternity" (1954, 95).

Such is the project of scientific Naturalism: to explain without loss, with the data science allows, all of reality, including mathematics, mind, and value.

Finally, the commitment to nature as science reveals it presents Naturalism with a dilemma. If Naturalism is committed to what current science tells us about reality, we have no reason to accept the current deliverances of science as definitive of reality. With respect to all of human history, the natural sciences are in their infancy, and contemporary natural science, for all we know, may turn out to be completely wrong. An assured commitment to the deliverances of contemporary science seems unfounded. If, on the other hand, Naturalism is committed to "the ideal science" – the so-called end of human inquiry about the nature of nature – then we simply have no idea what to believe right now. The ideal science, whatever that may be, offers no guidance in the present.

Strict and Broad Naturalism

Naturalism, understood metaphysically or ontologically, can be further divided between strict forms and more broad versions.

Strict Naturalists tend to privilege science in such a way as to downplay or explain away elements of our common understanding of human experience (sometimes condescendingly called "folk beliefs"). Paul Churchland (1998), for example, argues that the things we normally consider to be the basic elements of human cognition or mentality – thoughts,

beliefs, perceptions, desires, and preferences – are unnecessary to explain what happens in the physical world. Strict Naturalists believe that we can, without loss, reduce human mentality or consciousness to the physical; we are meat computers. They have also claimed that things such as free will, the self, and morality are illusions. It goes without saying that strict Naturalists think God is a folk belief that has been or can be "explained away." While folk beliefs may be useful for living our lives, they are not needed to explain the world we live in. In fact, proper scientific understanding explains such things away.

Broad Naturalists, while still metaphysical or ontological Naturalists, are more willing to accommodate at least some elements of our common or "folk" understanding of human experience. Broad Naturalists are more inclined to allow for things such as subjective experience, a self, free will, and morality. David Chalmers (1996), for example, argues that, despite their promise, Churchland-like scientific explanations fail to account for the reality of consciousness.[3] Chalmers is a Naturalist who thinks that human consciousness cannot be reduced to physical processes or properties. Broad Naturalists are motivated by the conviction that things like consciousness, morality, and freedom are irreducible to the physical.

Methodological Naturalism

Naturalism, as defined so far, talks about the *what* of reality. Methodological naturalism, on the other hand, tells us *how* we should conduct scientific inquiry: methodological naturalism forbids appeal to supernatural entities. Common features of a methodological naturalism are an emphasis on empirical research and a rejection of *a priori* theorizing (in which one starts by analyzing concepts and terms, apart from experience, and moves from this to claims about the world and what exists). Methodological naturalism, unlike Naturalism, is in principle neutral with respect to the existence of supernatural entities. So a scientist who is a methodological naturalist could be either a theist or an atheist. But methodological naturalists, atheist and theist alike, hold that science properly understood and practiced (and perhaps also philosophy) should not appeal to supernatural entities or forces.

Some Naturalists contend that methodological naturalism somehow leads to Naturalism. The story might go as follows: The naturalistic worldview has not been proven by a compelling argument (and probably never will), yet in science a naturalist worldview has proven enormously successful and is likely to continue to be explanatorily successful. Within that worldview there is no room for spooky supernatural powers likes gods and ghosts. The supernatural is scientifically irrelevant. Our current best understanding of the world needs no supernatural God. Likewise, progress in, say, neuroscience precludes our taking mental causation or mental entities as *real*: Platonic–Cartesian views of persons as composites of mind–body or spirit–soul–body are no longer scientifically tenable. The mind, which Gilbert Ryle derisively called "the ghost in the machine," is nothing more than the brain. Science has exorcised the world of gods and ghosts (even ghost-like minds). And if we don't need those entities to explain anything, why think that they exist? Good riddance.

But this "argument" is more hope than reality, and methodological naturalism makes no claims about domains outside of scientific inquiry.

3 Chalmers is not, to be clear, a mind–body dualist. He is a property dualist who argues that we must view consciousness as a fundamental natural property. Perhaps the difference is that broad Naturalists think that there are emergent levels in nature, whereas strict Naturalists think the lowest level, whatever it is, is the most/only real level.

The Historical Development of Naturalism

When we look at the historical development of Naturalism, it is important to keep in mind two motivations that reflect the methodological and ontological components mentioned earlier. The first is the attempt to do philosophy in a way that is consistent with the methods and results of science. The second is the attempt to understand the world without any appeal to supernatural beings or entities (God, souls, ghosts, etc.). These projects can be seen, to varying degrees, in the work of the many individuals who contributed to the development of contemporary Naturalism. Thus, it will come as no surprise that the development of modern science was crucial to the historical development of Naturalism.[4]

The scientific revolution of the 17th century transformed scientific inquiry. In the works of individuals such as Copernicus, Galileo, and Kepler we find an increased emphasis on experimental methods and a more extensive use of mathematics in the construction of theories, which were better at explaining the natural world than their predecessors. The success of these theories resulted in a spillover effect, as the emerging scientific methodology was applied to other domains.

This spillover can be seen in the work of several philosophers writing around this time. So, for example, we find the 16th-century philosopher Descartes attempting to model philosophical inquiry on geometrical reasoning by deducing conclusions from self-evident first principles that could be known outside of experience. Similarly, Thomas Hobbes brought the methodology of geometry to philosophy, in a slightly different way, attempting to synthesize and analyze the ideas that we obtain through sense experience. David Hume, generally considered one of the fathers of Naturalism, brought experimental methods to bear on many areas of inquiry. Evidence for this is seen in the subtitle to his *Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Sensory experiences, Hume claimed, provide the foundation for all areas of inquiry – not just science, but also logic, mathematics, morality, and politics.

Another important development in the history of naturalism involved several books written on the scientific method. Publications by the likes of John Herschel, John Stuart Mill, and William Whewell helped to strengthen the view that supernatural explanations did not belong in scientific work. They also helped to advance the growing interest in increasing the range of phenomena investigated through scientific methods. Yet, while scientists and philosophers from Copernicus and Galileo to Descartes and Herschel commended methodological naturalism, they were not Naturalists. Copernicus was a priest, Galileo wrote an essay on the proper interpretation of Scripture, Descartes offered several proofs of the existence of God, and Herschel was a devout Christian. How then did things progress from methodological naturalism to Naturalism?

The dramatic and unparalleled success of science in explaining more and more of the natural world contributed to the rise of Naturalism. In this regard, perhaps the most significant scientific contribution was Darwin's *Origin of Species* in 1859. Many have argued that Darwin's publication had monumental consequences for how we ought to view the world. Contemporary biologist Richard Dawkins, for example, writes, "although atheism might have been logically tenable before Darwin, Darwin made it possible to be an intellectually fulfilled atheist" (1986, 6). Prior to Darwin, design (by God) was considered necessary to

4 This section follows Rea (2004).

explain biological complexity; but after Darwin, divine design seemed unnecessary. Darwin's publication went a long way in boosting the explanatory power of nontheistic explanations of the natural world.

The significance of Darwin goes beyond the support it afforded atheism.[5] Prior to its publication, some areas of inquiry seemed shut off from scientific investigation. The most notable among these, of course, were questions concerning the origin of species, particularly of humans. Thus, as Michael Rea notes, before Darwin "there was still room for the idea that some domains of inquiry were better explored by other methods" (Rea 2004, 30). But Darwin's work opposed this idea. He suggested ways that science could investigate a whole host of phenomena, not solely the origin of species and life. Darwin's *Descent of Man* offered natural explanations of, among other previously untouchable items, morality and religion.

As post-Darwinian optimism for scientific investigation of these once sacred areas grew, Naturalism flourished. Science could explain everything, or so it seemed.

Naturalism has become an increasingly dominant view or approach to philosophy over the course of the last century. In John Dewey's early 20th-century work, *Experience and Nature* (1929), he attempted to achieve continuity of both methods and results between philosophy and the natural sciences. In the area of ethics, Dewey urged a "thoroughly empirical study" of the subject. In the second half of the 20th century, Harvard philosopher W.V.O. Quine did more than perhaps any other to advance a Naturalistic approach to philosophy and help bring it to dominance. Quine shared with Dewey the highest regard for the sciences and believed that philosophy should apply scientific methods to other areas of inquiry (Quine 1981). Unlike Dewey, however, Quine claimed that the subject matter of philosophy was continuous with science. For example, instead of worrying about the justification of beliefs (knowledge), Quine thought we should just study how the brain takes in sensations through the sensory organs and transforms them into beliefs (Quine 1969). The success of science, Quine thought, provided its own justification for the claims it made; it did not need outside justification from philosophy.

Following in the footsteps of these notable individuals, Naturalism and Naturalistic views have subsequently become dominant across a wide range of fields in philosophy. In the last several decades, there has been a flood of work under the banner of "Naturalism," with philosophers applying naturalistic approaches to virtually every area of human inquiry, including epistemology, rationality, jurisprudence, consciousness, ontology, and morality.

Arguing for Naturalism

Given this narrative of the rise of Naturalism, we can ask what are the motivations and arguments that have led to the dominance of Naturalism? The development and success of science may have caused or contributed to the acceptance of Naturalism, but the success of science is insufficient *reason* for accepting Naturalism. If one methodologically precludes the non-natural at the outset, it should be no surprise if the non-natural is then missing in one's explanations. Whether or not there could be arguments for naturalism is itself a matter of debate. As noted earlier, some take naturalism to be a research project; that is, a set of

5 Interestingly, Darwin himself was at varying times while developing and defending his theory a theist (Clark 2014, ch. 5).

methodological dispositions for how to approach inquiry. If this is the case, then these dispositions will determine what one counts as evidence, how one weighs different sorts of evidence, and so on. Thus, some think that naturalism itself isn't something that can be directly argued for.

Perhaps the most significant motivation for adopting a naturalist stance is a lack of belief in God. A uniting feature of naturalism, we noted, is a rejection of the supernatural. One might reject God's existence for a variety of reasons (the problem of evil, divine silence, lack of evidence). What is important is that once God falls out of the picture, Naturalism seems to many the only plausible alternative: Naturalism is the only game in town (Kitcher 1992; Stroud 1996).

Still, the plausibility of Naturalism involves more than arguments against the existence of God. A general feature of arguments in favor of Naturalism is an attempt to achieve a sort of simplicity in one's theorizing. Naturalists often argue that they can account for certain features of the world, like consciousness, the human person, or morality, without postulating the existence of non-natural entities, since such things are not explanatorily necessary. The motivation for this line of thought, which is commonly traced back to the medieval philosopher William Ockham and referred to as Ockham's razor or Occam's razor, is that we shouldn't multiply entities beyond necessity. That is, we should only postulate the existence of the bare minimum number of entities required for an adequate explanation.

Once one affirms Occam's razor, the crucial question is which entities we need and which we don't. Naturalists, across a wide array of areas, argue that supernatural entities are not (or are no longer) needed to explain our world. There might be such entities, but they are not necessary for belief. Moreover, once they are eliminated as explanatorily necessary, there is no reason to believe in them; we are neither required nor permitted to believe in them. Since they can adequately explain the world without appeals to the supernatural, Naturalists argue their position is rationally superior to non-naturalist philosophies.

Consider an analogy. Suppose you spend a week studying naturalism in a friend's upstairs study. Your friend has warned you that the attic is haunted by ghosts, which keep her awake at night and distracted during the day. As you are studying, you recall this and are pleasantly surprised by how quiet the room is. You stop every now and again and listen intently, but you hear no ghostly noises. After a few days, you find yourself coming to believe that there are no ghosts in the attic. You could be wrong, of course; those pesky ghosts could be sleeping or haunting someone else's house. So, while you concede the possibility of ghosts in the attic, you find yourself believing that there are none. For the Naturalist, belief in God may be like your belief in ghosts. It's logically possible that God, like ghosts, exists, but there's no reason to believe that either God or ghosts exist. And so you move on to your reading without ghost beliefs and to thinking philosophically without God beliefs. Just as one *could* believe in ghosts in the absence of evidence, so, too, one *could* believe in God in the absence of evidence. But why believe in either?

Naturalism and the Good

We don't have the space in this introduction to consider every topic involved in a complete discussion of naturalism. Let us consider just one area. For many religious believers, God is seen as the foundation of morality. If an all-powerful, all-loving, all-knowing, and perfectly good being exists, then it seems plausible that this being has authority over how we live our

lives. If God does not exist, as Naturalists believe, then we might wonder what grounds our moral claims and renders them authoritative. When we say that murder is wrong or honesty is good, are these claims true? And if these and other claims are true (as most believe), then what makes them so? Naturalists are faced with the challenge of providing answers to these questions, illustrating where ethics fits in with the natural world, and explaining how morality can exist in a world without God (or godlike beings, forces, or properties). Can our common understandings of morality fit with strictly Naturalist assumptions about the world?

Here's a way to put the problem. Richard Dawkins claims that the world that science discovers has "no design, no purpose, no evil and no good, nothing but blind, pitiless indifference." He goes on: "In a universe of blind physical forces and genetic replication, some people are going to get hurt, other people are going to get lucky, and you won't find any rhyme or reason in it, nor any justice. The universe we observe has precisely the properties we should expect if there is, at bottom, no design, no purpose, no evil and no good, nothing but a blind, pitiless indifference" (Dawkins 1995, 133). While the natural world is replete with elephants, asteroids, and atoms, which have extension, duration, and number, among many other properties, it is devoid of good and bad.

How can you get the Good from matter/energy in its various manifestations? Simon Blackburn argues that for a naturalistic ethics, "The problem is one of finding room for ethics, or placing ethics within the disenchanted, non-ethical order which we inhabit, and of which we are a part" (1998, 49). Naturalists have offered many responses to this not insignificant problem. We can divide these into those that attempt to preserve objective truth in morality and those that do not.

When we say that a moral claim is *objectively* true, we mean that what makes it true does not depend on our imparting or projecting moral features on the world in the way, say, that we impart the property *being worth five dollars* to a piece of paper.[6] For example, suppose Hitler had successfully united Europe around his Nazi ideals. Suppose further that, through indoctrination, all Europeans had come to believe that the elimination of Jews from Europe was a good thing. An objectivist would claim that, given such circumstances, all Europeans were simply wrong – they were wrong because their beliefs violated an objective moral standard, one which prohibited the killing of innocent human beings; moreover, those who killed Jews were similarly wrong, in spite of their feelings or desires or attitudes to the contrary.

One Naturalistic account that denies objective moral truths is error theory. Error theory holds that when we make moral claims, we are attempting to state truths (Mackie 1977; Joyce 2007). The error theorist goes on to argue that there are no such truths. Thus, all of our moral claims (that one should keep one's promises or that slavery is wrong) are in error. There is no truth when it comes to morality. Many find this position too extreme to accept.

According to expressivist accounts, very broadly speaking, when I say, "Murder is wrong," I am not claiming, as one might think, that there is a *fact* that murder is wrong. Rather, I am *expressing* my strong disapproval of murder. Expressivism is the view that moral claims aren't attempts to state propositions about the world, but are expressions of emotions, or approval or disapproval, or commitments we hold (Gibbard 1990; Blackburn 1993). Over the years, these accounts have grown in sophistication, but in expressivism's crudest form "Murder is wrong" could be translated into something like "Boo Murder!" and "Honesty is good" into something like "yay Honesty!" An alleged advantage of expressivist

6 See Shafer-Landau (2003) for an account of stance-independence and moral realism.

views over error theories is that, if they are correct, then when we engage in moral discourse we are not fundamentally in error, since we aren't attempting to state moral truths (Blackburn 1998). Thus, the fact that there are no such truths shouldn't concern us. However, contra expressivism, we do, or so it seems, make claims about facts when we make moral claims, and the lack of such facts should concern us. When we claim that injustice is wrong, we don't mean (merely) to express our own feelings or desires (or the feelings or desires of a group). Thus, it isn't clear how much of an advantage expressivism offers over error theory (Cuneo 2006; 2007; Egan 2007; Parfit 2011).

Conceding the inadequacy of these views, some Naturalists have attempted to retain the truth of some of our moral claims while accounting for and explaining them within a Naturalist worldview.[7] This falls under the category of ethical naturalism and is a species of moral realism: its proponents believe that there are objective moral facts and properties, so we aren't always in error when we make moral claims. We don't create the rules, so to speak. The rules stand over us and we are obliged to them; nevertheless, they are fully natural (Boyd 1988; Brink 1989; Copp 2003; Sturgeon 2006). Moral facts and properties, on this view, are *natural* facts and properties; that is, moral facts are or are reducible to the sorts of facts studied by the natural sciences.

How is it possible to understand moral facts and properties "natural as"?

Evolution and Ethics

One attempt to account for morality within a naturalistic framework, in a way that is continuous with science, has been to explain the origin and development of our moral judgments, values, and norms. Many factors are considered in this project, but our evolutionary history is claimed to have had a particularly large role in shaping morality. The well-known scientist E.O. Wilson declared that the time has come "for ethics to be temporarily removed from the hands of the philosophers and biologicized" (1975, 562).

At first glance, it makes sense to think of morality as at least partly continuous with the sciences. It is an empirical fact that humans make moral judgments, have certain feelings, and act in certain ways (often to obtain various ends or goals). Since these are empirical phenomena, we can look to science for a causal explanation for how they developed. For example, how is it that human beings have come to make moral judgments? This sort of explanatory project is different from the types of questions and concerns philosophers have traditionally had when they attempt to understand and account for morality. Philosophers traditionally have wanted to know if anything grounds the truth of moral judgments (if there are moral facts that make moral claims true). They have wanted to know the foundation of values and norms, not how human beings acquired the ability to make value judgments. Moreover, while humans may characteristically behave in various ways, ethicists are typically concerned with how humans *should* behave; they want to understand the sources of normativity, not the wellsprings of human action. So philosophers have usually focused on things such as the *justification* of moral claims, the *foundation* or *grounds* of morality or moral truth, and how we *ought* to live our lives. These questions seem to be of a very different sort than the empirical questions pursued by science. However, some have

7 Adequate space cannot be given to all the different metaethical positions naturalists have proposed. Other prominent positions include constructivist and relativistic accounts.

argued that the explanation of morality's origin and development has important implications for these philosophical questions (Street 2006).

At the outset, it might seem puzzling to propose that evolutionary forces had a role in the origin and development of morality. We might think that evolution, working through natural selection, would promote selfish behavior in individuals. How could evolution lead to the cooperative behavior that seems necessary for the development of morality or the altruistic behavior that comes at the expense of an individual's own reproductive fitness? How do we get from "selfish genes" to nonselfish behavior? As it turns out, a good case can be made that evolutionary forces played some role in the development of cooperative or helping behaviors for many species, including *Homo sapiens*. To see how this could work, consider kin selection. A good example of this process comes from social insect colonies, such as the *Hymenoptera* (bees, ants, and wasps). In these colonies, we find high degrees of cooperation. Labor, food, and information are shared, and individuals not only cooperate but also sometimes sacrifice their lives in the colony's defense. What explains how evolution could lead to this cooperation and self-sacrifice? The answer lies in the fact that the colony is essentially a big family.

One common way of improving an organism's gene ratios – the way most of us are familiar with – is through reproductive success, promoting copies of the gene in more organisms in succeeding generations. However, gene ratios can also be improved if "sacrificial behavior" permits an organism's kin to have greater reproductive success. This happens because an organism and its relatives share many of the same genes. Kin selection can explain how social insects such as worker bees, which share virtually identical genes, were able to evolve apparently "selfless" traits that aid the survival and reproduction of the queen bee. By doing so, they were able to promote the very similar genes of their kin.

Nevertheless, helping one's kin and even engaging in helping behavior are not the same as being moral (Joyce 2005; Kitcher 2006). And apparently selfless traits are not the same as moral virtues.[8] Under most construals, morality occasionally requires helping non-kin or helping another with the motivation to benefit that other (and not, say, with the hidden motivation to spread one's genes into the next generation). Moreover, morality seems to have a kind of normative force or authority that scientific descriptions of the development of kin-regarding behaviors omit. One could help kin and have no thoughts about what one ought to do. Moral behavior, however, requires thoughts of this sort (or at least the capacity to act from such thoughts). There is much more to be said here, but the problems of establishing morality on the basis of kin selection are clear.[9]

While some have just been concerned with empirical questions – examining the ways in which evolutionary forces might have led to the development of cooperative behavior – others have gone further, to explore the significance this influence might have for philosophical questions regarding the existence and nature of morality. Some have argued that these evolutionary influences suggest ways we should approach ethical theories of how we ought to live our lives.

Rosalind Hursthouse (1999) grounds morality in a biologically informed account of human nature, one stripped of the decidedly non-natural *telos* of Aristotle. Aristotle's

8 Some argue that morality is not just concerned with our emotions and behaviors, but, crucially, with moral *judgments* about how we ought to live our lives. If that means that morality requires us to make moral judgments before we can act morally, then their conception of morality seems too intellectualistic.

9 We are considering just a few of the issues involved in treating ethics as continuous with science; in this case, evolutionary biology. We leave aside other evolutionary explanations of morality.

ideal human (the human *telos*) is beyond nature (no one, for example, has seen or even could see, touch, taste, smell, or hear this perfected human). Hursthouse claims that the human standard of goodness is grounded in species-typical goods such as survival and reproduction, but also, along with higher species, in freedom from pain and the enjoyment of pleasure. Finally, since humans are by nature social animals, their good is partly constituted by their contribution to a fully functioning group. Hursthouse contends that social goods, evaluated and affirmed by a robust conception of human rationality, ground what we mean when we way say that a human being is good. In Chapter 2 of this volume, Flanagan et al. argue along similar lines: "One common rationale for favoring a norm or set of norms is that they are suited to modify, suppress, transform, or amplify some characteristic or capacity belonging to our nature – either our animal nature or our nature as socially situated beings" (pp. 29).[10] As self-reflective, social animals we have come to affirm certain values as essential to a life well lived, including happiness, love, friendship, and respect.

It is an open question, though, how well-grounded these values can be without some idea of a perfected human nature (a *telos*) to determine which of our many and conflicting desires and emotions are morally praiseworthy and which are not. But the idea of a perfected human nature lies outside of nature and so is not available to the Naturalist. The problem is that we find in nature many types of species-typical "goods" that are morally abhorrent: ants that take other ants as slaves, or sea otters that rape baby seals. Humans, in addition to caring for others and desiring friendship, are also inclined toward selfishness and tribalism (and for good evolutionary reasons). So, we cannot simply look to nature for a guide to living our lives. Without an independent, perhaps non-natural standard of the good, how can we demonstrate that ignoble desires are bad and noble ones are good?[11]

Finally, an evolutionary justification of morality in terms of its instrumental or practical value may not be the type of justification or grounding most of us are seeking. Many people believe that morality makes certain demands on us; it has, as Richard Joyce (2005) puts it, "clout." Instrumental or practical justifications, so it seems, cannot ground or justify the authority or demands of morality. Even if morality has practical value, why should I, in any particular circumstance, think it applies to or has authority over me?

For Joyce, the problem is exacerbated. Joyce contends that the evolutionary origins of morality debunk the claim that there are objective moral facts. He contends that since we can provide an evolutionary account of the development of our moral concepts without any appeal to "moral facts," these facts are unnecessary to an adequate explanation and, thus,

10 Some argue that psychological studies can help us determine where emotional responses might be distorting our moral response to certain thought experiments or situations. After doing so, we can note ways in which evolutionary influences might have led to these emotional responses, and, subsequently, come to a better normative ethical theory in light of this information. Joshua Greene (2003) uses this sort of argument to argue for consequentialism over other normative ethical theories.

11 Robert Richards (1987) argues that evolution has constructed human beings to act for the community good. The "community good," he contends, is what we mean by "moral." And since human beings are moral beings – that is, since this is an unavoidable condition produced by evolution – each of us ought to act for the community good. There are several problems with this argument. For one, it is doubtful that "the good of the community" is what we mean by the word "moral." But, beyond this, Richard's conclusion – that each of us ought to act for the community good – does not seem to follow from his argument. Why must we all act for the community good? Richards seems to be moving from an "is" to an "ought."

should be rejected.[12] However, Joyce thinks that since rejecting moral facts would undermine morality, we should treat morality as a "useful fiction."

For the naturalist, making sure that our accounts of morality are consistent with our best science is not optional. But it requires answering some profound questions: Is it possible to base value on fact (morality on, say, evolutionary forces)? Is morality thus construed as true or is it a fiction (useful or otherwise)? Does a naturalistic account of morality require substantial revision to our traditional understandings of morality and its demands?

Discontents

We have briefly surveyed the problems and prospects of a Naturalistic account of ethics. There are, as we've mentioned, a great many other issues that require adequate Naturalistic explanations. Peter Railton, for example, raises the problem of a naturalistic account of free will: "Where in the portrait of the natural world and its laws, and of the human organism and its physio-chemical nature, does one find anything corresponding to *free will* as we understand it?" (2012, 22). In *Darwin's Dangerous Idea,* Daniel Dennett (1995) claims that evolution is a universal acid that eats through (undermines or debunks) many of our very ordinary ways of understanding the world; while God is the first to go, human uniqueness, morality, freedom, the soul, an enduring self, and consciousness succumb to evolution's corrosive powers. Dennett, following Jeremy Bentham, calls human rights, "nonsense on stilts" (but, inexplicably, "good nonsense").

Some critics reject Naturalism for its inability to find a place for morality, freedom, and the self (as just three examples). These critiques commonly claim that strictly Naturalistic assumptions cannot adequately explain what the Naturalist hopes to explain. Timothy Williamson (2011), a self-professed atheist, rejects Naturalism, especially of the scientific variety, because of its inadequacy at accounting for everything: history, literature, logic, linguistics, and mathematics. As noted, strict Naturalists are criticized for failing to adequately explain subjective conscious experience (say, the sweet taste of honey or the sharp feel of pain); despite the promise of a science of the mind, not a single item of conscious experience has been shown to be reducible to the brain and its chemistry. Some strict Naturalists have gone so far as to deny the existence of subjective consciousness, and even the continued identity of persons over time. But consciousness and persons are, at least *prima facie*, features of the natural world that need to be explained. Failure to do so, critics argue, represents a failure of Naturalism and provides good reason to reject it.

Naturalism's alleged failures are, the committed Naturalist hopes, its opportunities. Like the belief that roses are red (that colors are in the objects they represent), some cherished and commonsensical beliefs will be scientifically explained away. It is an open question, however, whether or not free will and morality, for example, have been or can be scientifically dispatched. And, like heliocentrism and evolution, some philosophical beliefs may eventually yield to scientific explanation (without science having explained them away). Like the science it rightly esteems, Naturalism is in its infancy, and what it can or cannot explain remains to be discovered.

12 Sharon Street (2006) offers a slightly different argument for the same conclusion. She claims that because evolutionary forces are not truth-directed, but rather push toward reproductive success, we cannot be confident that any of our moral beliefs are true.

The Shape of the Book

This collection contains mostly adherents but also some critics. Naturalism is part of an ongoing and fundamental debate in contemporary philosophy. To present it without criticism seemed imprudent. Each defender of Naturalism has been asked to define Naturalism as they understand it and then to explore this or that area of philosophy from the perspective of naturalism. Critics, no surprise here, were asked to define and then criticize various naturalist proposals. We offer definitions, defenses, and criticisms of naturalism *simpliciter*, as well as developments of a naturalistic understanding of various philosophical domains: science, metaphysics, mind, knowledge, truth, mathematics, free will, ethics, religion, and social/political philosophy. Since discussions of naturalism/non-naturalism have been so lively in ethics, this was afforded some prominence.[13]

Unlike "the assured results of modern science," there are no assured results of Naturalism in any of its guises. Indeed, there is no definitive understanding of Naturalism itself. Defined too narrowly, it leaves out wide swaths of human thought and experience; defined too broadly, it includes many of those things that naturalists hope to exclude. Nonetheless, we hope the reader gets some sense of philosophy as exploration, as trying out this or that route, all the while lacking a clear map or even a sense of destination. Most of the authors, like most philosophers, have taken their path, but they have not taken that path blindly, and even their failures are illuminating. We hope the reader catches the sense of excitement that philosophers feel as we explore, and a sense of intellectual humility as we attempt to see through the glass, darkly.

References

Blackburn, S. (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Blackburn, S. (1998). How to Be an Ethical Anti-Realist. *Midwest Studies in Philosophy* 12: 361–375.

Boyd, R. (1988). How to Be a Moral Realist. In *Essays in Moral Realism*, edited by Geoffrey Sayre-McCord, pp. 187–228. Ithaca: Cornell University Press.

Brink, D.O. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Chalmers, D., Manley D.M. and Wasserman, R. (eds.) (2009). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

Churchland, P.M. (1988). *Matter and Consciousness*. Cambridge, MA: MIT Press.

Clark, K. (2014). *Religion and the Sciences of Origins*. London: Palgrave-Macmillan.

Copp, D. (2003). Why Naturalism? *Ethical Theory and Moral Practice* 6: 179–200.

Cuneo, T. (2006). Saying What We Mean: An Argument against Expressivism. In *Oxford Studies in Metaethics*, vol. 1, edited by Russ Shafer-Landau, pp. 35–71. Oxford: Oxford University Press.

Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

13 In case you are thinking, "Why wasn't this or that (obviously better) topic discussed?", the list of topics was developed in conversation both with Wiley-Blackwell's editors and with the contributors themselves. Moreover, Wiley-Blackwell imposed word limits. Given that we couldn't discuss everything, we did our best to select topics that most deeply manifested the prospects and problems of Naturalism.

Dawkins, R. (1986). *The Blind Watchmaker*. New York: Norton & Company.
Dawkins, R. (1995). *River Out of Eden*. New York: Basic Books.
Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.
Dewey, J. (1929). *Experience and Nature*. London: George Allen and Unwin.
Egan, A. (2007). Quasi-Realism and Fundamental Moral Error. *Australasian Journal of Philosophy* 85(2): 205–219.
Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press.
Greene, J. (2003). From Neural "Is" to Moral "Ought": What Are the Moral Implications of Neuroscientific Moral Psychology? *Nature Reviews Neuroscience* 4: 847–850.
Hursthouse, R. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.
Joyce, R. (2005). *The Evolution of Morality*. Cambridge, MA: MIT Press.
Joyce, R. (2007). *The Myth of Morality*. Cambridge: Cambridge University Press.
Kitcher, P. (1992). The Naturalists Return. *Philosophical Review* 101.1: 53–114.
Kitcher, P. (2006). Ethics and Evolution: How to Get Here from There. In *Primates and Philosophers,* Frans de Waal, pp. 120–139. Princeton: Princeton Science Library.
Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin Books.
Parfit, D. (2011). *On What Matters*. Oxford: Oxford University Press.
Quine, W.V.O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Quine, W.V. (1981). *Theories and Things*. Cambridge, MA: Harvard University Press.
Railton, P. (2012). Nietzsche's Normative Theory? The Art and Skill of Living Well. In *Nietzsche, Naturalism and Normativity*, edited by Christopher Janaway and Simon Robertson. Oxford: Oxford University Press.
Rea, M. (2004). *World Without Design: The Ontological Consequences of Naturalism*. Oxford: Oxford University Press.
Richards, R. (1987). *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press.
Rosenberg, A. (2011). *The Atheist's Guide to Reality: Enjoying Life without Illusions*. New York: W.W. Norton.
Ross, D., Ladyman, J., and Kincaid, H. (2013). *Scientific Metaphysics*. Oxford: Oxford University Press.
Schroedinger, E. (1954). *Nature and the Greeks and Science and Humanism*. New York: Cambridge University Press.
Sellars, R.W. (1922). *Evolutionary Naturalism*. Chicago and London: Open Court.
Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge & Kegan Paul and New York: The Humanities Press.
Shafer-Landau, R. (2003). *Moral Realism: A Defense*. New York: Oxford University Press.
Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127: 109–166.
Stroud, B. (1996). The Charm of Naturalism. *Proceedings and Addresses of the American Philosophical Society* 70: 43–55.
Sturgeon, N. (2006). Ethical Naturalism. In *The Oxford Handbook of Ethical Theory*, edited by David Copp, pp. 91–121. Oxford: Oxford University Press.
Williamson, T. (2011). What is Naturalism? *New York Times*, September 4. Available from: http://opinionator.blogs.nytimes.com/2011/09/04/what-is-naturalism/ (last accessed July 15, 2015).
Wilson, E.O. (1975). *Sociobiology*. Cambridge, MA: Harvard University Press.

2

Naturalizing Ethics

OWEN FLANAGAN, HAGOP SARKISSIAN, AND DAVID WONG

Introduction

In this chapter we provide (1) an argument for why ethics should be naturalized, (2) an analysis of why it is not yet naturalized, (3) a defense of ethical naturalism against two fallacies – Hume and Moore's – that it allegedly commits, and (4) a proposal that normative ethics is best conceived as part of *human ecology* committed to *pluralistic relativism* (Flanagan 1995; 2002; Wong 1984; 1996; 2006b). The latter substantive view, supported by a neocompatibilist view of human agency, constitutes the essence of *Duke naturalism*. It provides a credible substantive alternative to bald or eliminativist *Australian ethical naturalism*, especially one that supports moral skepticism (Mackie), and to the more reticent *Pittsburgh naturalism*.[1]

Naturalism in the Broad Sense

Ethical naturalism is a variety of a broader philosophical naturalism, so it will be good to say what naturalism in the broad sense is. According to the *OED*, the original *philosophical* meaning of the term "naturalism" dates back to the 17th century and meant "a view of the world, and of man's relation to it, in which only the operation of natural (as opposed to supernatural or spiritual) laws and forces is admitted or assumed."

1 McDowell coined the term "bald naturalism" (McDowell 1996) and sometimes characterizes it in a way that engenders or is akin to moral skepticism. In principle, a naturalist might be a moral skeptic, believing that there are no moral properties as ordinarily conceived and thus that moral propositions are literally false (or meaningless). There is nothing ethics is about, and so forth. See Mackie (1977). For a more recent treatment, see Joyce (2001). *Duke naturalism* is also superior to *Rutgers naturalism*, but showing this will await a subsequent paper. It is possible that *Duke naturalism* and *Michigan naturalism* can be coalesced into *Duke–Michigan naturalism*.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

In a presidential address to the American Philosophical Association, Barry Stroud writes:

> Naturalism on any reading is opposed to supernaturalism…By "supernaturalism" I mean the invocation of an agent or force which somehow stands outside the familiar natural world and so whose doings cannot be understood as part of it. Most metaphysical systems of the past included some such agent. A naturalist conception of the world would be opposed to all of them. (Stroud 1996)[2]

Stroud's comment can either be about ontology (i.e., that naturalists reject agents or forces that stand outside the natural world) or be about methodology (i.e., that naturalists reject the invocation of such agents or forces in their philosophical projects). These issues will be fleshed out later in the chapter. For now, we can note that ethical naturalism holds a number of thin ontological commitments, and some more substantial methodological ones.

Why Ethics Isn't Naturalized

Ethical naturalism has a fair number of philosophical advocates, but most people reject it – including many in the academy. The reason has to do with the close connection in American culture (and throughout much of the world) between ethics and religion, and thus with the supernatural. Over 90% of Americans believe in a personal God who answers their prayers. Similar numbers believe in heaven (slightly fewer believe in hell). They believe that God, Yahweh, Allah is the source of moral law and that He rewards (or punishes) "souls" based on how well they conform to the moral law. Most people believe in God and think that moral knowledge is knowledge of what God creates or endorses as "good," "bad," "right," and "wrong." Religious reasons are offered and accepted in America, and most of the rest of the world, as legitimate moral grounding reasons. Ethical naturalism, as a species of naturalism, rejects religious grounding reasons.

Let us call an individual a *scientific naturalist* if she does not permit the invocation of supernatural forces in understanding, explaining, and accounting for what happens in *this* world. An *ethical naturalist* (assuming this person already accepts scientific naturalism) applies the same principled restriction to describing, explaining, recommending, endorsing, prohibiting, and justifying values, norms, actions, principles, and so on.[3] In other words, the complete warrant for any norm or value must be cashed out without invoking the views or commands of a divinity.

Call the folk view that there is a creator God, but one to whose thoughts we have no access, and who does no work in this world, has no effect on how this world operates, "folk naturalism." This view is relatively common. Indeed, it must be if we are to make sense charitably of the behavior of scientists who restrict themselves to physical explanation of physical phenomena but believe nonetheless in God.

2 For other discussions of naturalism in philosophy, see Bouwsma (1948), Kitcher (1992), and Rosenberg (1996), as well as entries on "Naturalism" in Honderich (1995) and Audi (1999).

3 Not all scientific naturalists are ethical naturalists, but it would be a rare bird who was an ethical naturalist but not also a scientific naturalist. Typically, a view that incorporates these two kinds of naturalism is *epistemically imperialistic* as far as *this* world goes.

Most people accept that science can legitimately take the folk naturalistic stance. Seamstresses, carpenters, plumbers, auto mechanics, and scientists all practice as if they are committed to folk naturalism. Scientists take official vows, as it were, when they declare themselves to be scientists in some domain of inquiry. There is no glaring inconsistency in thinking that God set up the world in such a way that scientists or auto mechanics can describe, explain, predict, and manipulate what happens in it. The fact that there is no inconsistency does not mean it makes epistemic sense – that there are any good reasons – to believe in God.

When it comes to ethics, though, most people will balk at restricting themselves to folk naturalism. For complex reasons (though having to do in some large measure with the importance of morality), most people would like to see moral value justified in a very strong way. It would be good if moral values, beliefs, norms, and the like had something like the necessity that mathematical theorems have. One way this could work would be *if* an omniscient and all-loving being made the rules and then provided us with epistemic access to them. There are various familiar stories of how this works. Human souls exist prior to bodily implantation in the company of "The Good" and, once embodied, have the ability to remember, recollect, or intuit what is good, bad, right, and wrong (Plato; and with certain modifications, G.E. Moore). God directly illuminates faithful human hearts and minds through grace (Augustine). God produces a world and mind such that His perfect nature can be deductively established and, with some additional difficulty, His will can be known (Anselm, Descartes, Alvin Plantinga).[4] God speaks to certain sages, who write down His moral rules in sacred texts (the Torah, the Old and New Testaments, the Qur'an), and so on. The last view – that we have epistemic access to God's will and commands through sacred texts – is the dominant view: moral values, norms, and principles have their ultimate ground or warrant in divine revelation. There are many good values expressed in these sacred texts – the Golden Rule, for example. Most naturalists will endorse the Golden Rule, but all will reject that its warrant relies on any supernatural source.

Both folk naturalism and scientific naturalism are methodological and domain-restricted. Neither view *warrants* belief in any supernatural forces or entities outside this world, but both *allow* it. Naturalism in the broad sense does not prohibit there being supernatural forces or beings; it is just that such forces or beings do not – at least since she's been up and running – have causal (or any other sort of) intercourse with Mother Nature. A stronger form of naturalism says that what there is, and all there is, in this and any *actual* world is natural. Stronger still would be the claim that what there is, and all there is, in any *possible* world is natural – that it is impossible for there to be *any* world that contains supernatural beings, entities, and the like. (This view has its attractions because it has rhetorical

4 Kitcher (1992) and Stroud (1996) claim that naturalism is pretty much the only game in town. This claim has credibility to the degree that they intend some version of the idea that for the purposes of doing ontology (or for naturalized epistemology or ethics), divine agency does not need to be introduced to play an explanatory role. Stroud, for example, thinks that Plantinga is a naturalist when it comes to descriptive epistemology, but not when it comes to normative epistemology. However, there are other major contemporary philosophers whose views should also give him pause, such as W.P. Alston, Alasdair MacIntyre, and Charles Taylor. MacIntyre (1988; 1991) holds a two-level view; like Aquinas, he thinks that there are natural and supernatural justifications of norms. The natural justifications are satisfactory but the divine ones are "more" ultimate. Taylor (1989) expresses the idea that perhaps God can play a role in the justification of our ethical norms.

force in telling everyday supernaturalists that one is just not going to yield them any ground. However, it epistemically overreaches.)[5]

At any rate, what seems warranted is this: there are no good epistemic reasons to believe that there are any of the entities, processes, or forces of the sort posited by any supernaturalist ontology. Call this *quietistic ontological naturalism*. The view is bold but quietistic at the same time – thus its name. The Buddha, at his inaugural address, claimed that no human, himself included – as enlightened as he was – was in any position to give epistemically respectable answers (possibly to even formulate epistemically respectable questions) on matters such as those on which both ordinary religious folk and woolly metaphysicians are inclined to speak. The Buddha was a Wittgensteinian as far as epistemology goes: "Whereof one cannot speak, one ought to be silent." For present purposes, consider us committed to this form of naturalism. What is warranted, all things considered, is a form of ontological naturalism about *this* world, which is the only world we have reason to believe exists. Thus, for all we know, what there is – and all there is – is the natural world. Because the conception of what is "natural" is not fixed, the central concept in the motto lacks a clear and determinate meaning. Still, vague as it is, the view is not friendly to theism.[6]

5 *Ontological naturalism* is the view that what there is, and all there is, is natural. Everything that exists/has existed, happens/will ever happen, is a natural phenomenon, process, or event. Every property, event, process, and thing, if it *genuinely* exists/is happening, did exist/happen, or will exist/happen, is natural. This is the right definition, but it is not especially helpful until we specify what it means to be "natural." This is surprisingly hard. Imagine a world governed by Newtonian physics and Darwin's theory, supplemented by whatever chemistry and molecular biology (etc.) go neatly with them. In this world, "natural" would mean something like this: what there is and all there is, is whatever Newtonian physics and the principles of Darwinian biology say there is or can be. Furthermore, all the events in this world are explained, or explainable, by the causal laws of (or extendable from) these two theories (plus chemistry, etc.). This much would give us the familiar picture of the natural as comprising law-governed material, and it would provide us with a story of why the physical universe contains what it does and behaves as it does, as well as of how life came to be, evolved, and so on. The trouble is that contemporary physics is wilder and woollier than Newton's physics. There is quantum indeterminacy. There are particles or strings that are the same size in relation to a single proton as a single proton is in relation to the known universe. There are (*if* string theory is true) at least ten spatial dimensions, as well as one for time. Gravity is weak because gravitons escape from our universe into other ones through wormholes in space.

The story that our universe originated in a Big Bang 13 or 14 billion years ago was always somewhat mysterious. How did the singularity that banged get there? Some will say that there was "no there" then. Well, there was no "then" either because space and time as we know them only came into being when the "thing that wasn't quite there then" banged. Many sensible people have found this explanation less than satisfying. As we've said, more than 90% of Americans believe that a supernatural force, indeed a personal God, created the universe. If scientists can get away with postulating that the singularity that was not really "there then" led to there being "hereness," "thereness," "nowness," "thenness," and so on, many theists feel licensed to posit their own kind of mystery. See Craig and Sinnott-Armstrong (2003).

Contemporary physics now seriously toys with explaining away the mystery of the Big Bang in this way: our universe, which appears to be four-dimensional (three space, one time), isn't. The thing that appears that way is really eleven-dimensional, and it was formed – guess when? – 13 or 14 billion years ago, when a preexisting universe wormed its way into empty space in our vicinity. How did that sneaky universe get going? Same way. It's universes all the way back, down, and so on. See, for example, Greene (2000).

6 The epistemological humility called for is not so humble that it tolerates agnosticism. The agnostic thinks it makes epistemic sense to stop with three possibilities: theism, agnosticism, and atheism. But the quietist thinks the agnostic has been tricked into playing a fool's game. There is no reason to play the game or address the questions that force the familiar three choices. Just say "no" to talking about the supernatural and only then, apparently, will you see that nothing epistemically respectable can be said on such matters.

Why Naturalize?

The ontological naturalism advanced in this chapter opposes belief in supernatural forces. It also rejects various forms of ontological dualism, such as a sharp bifurcation of humans and the rest of the natural world, as well as mind–body dualism.

Ethical naturalism is not chiefly concerned with ontology, but with the proper way of approaching moral inquiry. Ethical naturalism thus has a number of methodological commitments, only some of which consist in a rejection of supernatural forces when explaining or justifying values and principles.[7] If naturalism were only opposed to supernaturalism, then the category of ethical naturalists would be overly inclusive, and if it were only committed to being receptive to findings from the natural sciences, if its most plausible core doctrine was a kind of open-mindedness, as Barry Stroud (1996) has suggested, then it would be difficult to see how it could represent a distinctive view.

However, naturalistic ethics does have a number of substantive methodological commitments. Chief among these is the belief that moral philosophy should not employ a distinctive *a priori* method of yielding substantive, self-evident, and foundational truths from pure conceptual analysis. The claims of ethical naturalism cannot be shielded from empirical testing. Indeed, the naturalist is committed to there being no sharp distinction between her investigation and those of relevant other disciplines (particularly between epistemology and psychology). In other words, ethical science must be continuous with other sciences.

In order to better understand what naturalistic ethics entails, it might be helpful to consider varieties of moral theory that are not supernatural, yet not natural either. Some such theories are semantic and maintain that moral terms (or predicates) cannot be cashed out using non-normative terms (or predicates). More frequently, such theories affirm a metaphysical thesis that naturalists deny – namely, the existence of irreducible and non-natural moral facts or properties.[8] Other non-naturalists maintain the autonomy of the moral: that morality is essentially autonomous from other forms of inquiry, namely, from the natural sciences.

Consider, for example, morality as conceived by Immanuel Kant. In the *Groundwork*, Kant writes that a "worse service cannot be rendered morality than that an attempt be made to derive it from examples." Trying to derive ethical principles "from the disgusting mishmash" of psychological, sociological, or anthropological observation, from the insights about human nature that abound "in the chit-chat of daily life" and that delight "the multitude," and upon which "the empty headed regale themselves," is not the right way to do moral philosophy (Kant 1785/1964).

7 See also Putnam (2004).

8 Indeed, Pigden (1991) suggests that the belief in the nonexistence of such irreducible *sui generis* moral properties or facts is what unites the category of naturalists. Most naturalists maintain that if ethics provides a respectable kind of wisdom, then moral properties, values, virtues, norms, oughts, and so on must be analyzed in ways that do not involve ontologically queer properties, forces, or commitments. The kind of naturalism defended here is quietistic with regard to the existence of supernatural forces or entities. This is also true of natural moral properties. We imply no position on the question of whether there really are, or are not, moral properties in the universe in the sense debated by moral realists, antirealists, and quasi-realists. The important thing is that moral claims can be rationally supported, not that all the constituents of such claims refer (or fail to refer) to "real" things. Furthermore, in both the realism/antirealism cases and the cognitivist/noncognitivist case, different answers might be given at the descriptive and normative levels. Mackie (1977) is an example of a philosopher who thought that ordinary people were committed to a form of realism about values but were wrong. In spite of this, Mackie saw no problem with advocating utilitarianism as the best moral theory, and in that sense was a cognitivist – a cognitivist antirealist, as it were.

What is the right way to do moral philosophy? According to Kant, we need "a completely isolated metaphysics of morals," a pure ethics unmixed with the empirical study of human nature. Once moral philosophy has derived the principles that ought to govern the wills of all rational beings, then and only then should we seek "the extremely rare merit of a truly philosophical popularity."[9]

Kantian ethics, *qua* philosophical theory, might not seem to be openly supernaturalistic. However, it is not naturalistic either. Kant maintains, for example, that postulating the existence of God is essential to ethics.[10] What's more, he claims that the self, the will, and the laws of freedom reside in the realm of the noumena – which is disconnected from all phenomena that could be studied by science. Kant's ethics cannot be naturalistic because we cannot give a naturalistic account of these things. For example, we cannot account for a faculty of pure practical reason that possesses moral principles not gleaned from observation and assessment of human practices that work differentially well to meet our aims, and which, in addition, fits with the findings of the mental sciences. There is no such faculty that meets these criteria, and thus no faculty to account for.

Kant's rationale is transcendental. Ethical naturalism is nontranscendental, since it rejects divine command or, in this case, since it will not locate the rationale for moral claims in the *a priori* dictates of a faculty of pure practical reason. Thus, ethical naturalists will need to explain the appeal of transcendental rationales and explain why they are less credible than pragmatic rationales (possibly because they are disguised forms of pragmatic rationales). Suppose, as seems plausible, that Kant intended his *grounding* of the categorical imperative in pure practical reason to both rationalize the categorical imperative and motivate us to abide by it. If one denies (as we do) that there is such a thing as pure practical reason, and if one also thinks that the categorical imperative expresses deep moral insight, then one needs to give an alternative account of the deep insights he expressed.

Unless one is an eliminativist or a physicalist in the reductive sense – that is, a bald naturalist – then reasons exist, as do norms and ideals. Reasons, furthermore, can be causes. However, being *a reason that causes* is not the same as being *a reasonable cause*; a motivating reason is not, in virtue of being motivating, something that it is reasonable to believe in or invoke in justification of one's (other) thoughts or actions. The belief that "Santa Claus will deliver coal to me unless I behave myself" will motivate, but it is not the sort of thing we think a sensible adult should believe in, let alone be motivated by.

However, since beliefs that have contents that don't refer are no problem for the naturalist, the causal power and efficacy of beliefs about things that don't exist are not things that worry the naturalist, either. It is largely a matter of psychological, sociological, and anthropological inquiry why different sorts of things are motivating reasons; that is, why certain reasons and not others motivate at different times and places. The task for the normative

9 Thanks (or no thanks) to Kant, the dominant conception of the intellectual division of labor makes a sharp distinction between moral philosophy and moral psychology. Moral philosophy is in the business of saying what ought to be, what is really right and wrong, good and evil, what the proper moral principles and rules are, what counts as genuine moral motivation, and what types of persons count as genuinely good. Most importantly, the job of moral philosophy is to provide philosophical justification for its "shoulds" and "oughts," for its principles and its rules.

10 Despite his commitment to the project of the Enlightenment, Kant in fact believed in God – namely, the God of pietistic Lutheranism. And he believed that God was, in fact, the ultimate source of morality.

naturalist is to recommend ways of finding *good* reasons for belief and action and to indicate why it makes sense to be motivated by such reasons.[11]

Just as a naturalist cannot accept the postulate of a faculty of pure practical reason, she also cannot accept the notion (found in Kant) that humans have metaphysical freedom of the will. Descartes famously articulated the idea this way: "But the will is so free in its nature, that it can never be constrained…And the whole action of the soul consists in this, that solely because it desires something, it causes a little gland to which it is closely united to move in a way requisite to produce the effect which relates to this desire" (1649/1968). The 20th-century philosopher Roderick Chisholm puts the point about free agency, what he calls "agent causation," thus: "[I]f we are responsible…then we have a prerogative which some would attribute only to God: each of us when we act, is a prime mover unmoved. In doing what we do, we cause certain things to happen, and nothing – or no one – causes us to cause those events to happen" (1966).[12] Descartes' and Chisholm's views are openly non-naturalistic. This sort of free will violates the basic laws of science, so the naturalist must offer a different analysis.

The most plausible view, "neocompatibilism," holds that free will is compatible with determinism (Flanagan 2002). A neocompatibilist rejects the dialectic that frames the problem as one between free will and determinism because our best physics now says that there are both ontologically deterministic and indeterministic causal processes, so determinism is not the issue. Causation is. And no matter how one views things, causation is ubiquitous. So far, it does not look as if indeterministic processes at the quantum level "percolate up" to macro levels. However, in principle, they might. If they do, some macroprocesses might not be deterministic. But it is a nonstarter to think this will help secure a place for "metaphysical freedom of the will" or "agent causation." Consider the hypothesis that there are in fact quantum-gravitational effects or processes in the microtubules of certain neuronal segments. Such random swerves do nothing to secure anything like agent causation, which involves an agent doing something as a prime mover himself unmoved.

On the other side of the dichotomy sits the concept of "free will." Naturalists from Hume on have tried to tame the concept of free will. The picture that Hume and other compatibilists paint is a fairly good fit with our best contemporary science. However, to really make the compatibilist position work, one will need to read Hume et al. as resisting the posit of a distinctive faculty of will; that is, as rejecting the faculty psychology within which free will, reason, imagination, and their suite historically are situated.

What is "new" – and thus what warrants the name "neocompatibilism" – is the outright denial of any faculty that fits the description of free will. Why? Because the concept utterly fails to locate anything significant that we mean to be talking about.

11 Motivational grounds to one side, there is always the interesting question of whether, even if we judge the reasons motivating some norm(s) unwarranted, we judge the norm(s) themselves unwarranted. There is no strict implication. One is inclined to say that even if you behave well only because you believe "Santa Claus thoughts," you should still behave well even though there is no Santa Claus – albeit for non-Santa Clausy reasons. On the other hand, it just may be the case that, across multifarious social contexts, things like "Santa Claus thoughts" motivate as well – if not better – than "Mom and Dad disapprove of" thoughts. If this is so, we need an explanation of how beliefs in certain kinds of nonexistent objects can motivate, and motivate powerfully. False beliefs that produce goods are an interesting phenomenon, but they create no special problem for the naturalist.

12 Chisholm might, however, say that the self depends on and is often affected by the body, and so is part of the natural world in this limited sense.

There is no such thing as "will," and thus no such thing as a "free will." (Here is one of the rare places where the eliminativist move is totally warranted.) There is no faculty of will in the human mind/brain. Talk of dedicated faculties can be useful when speaking of sensory and perceptual systems, but no respectable cognitive neuroscientist thinks there are distinctive faculties of will, reason, imagination, and the like.[13] If there is no such faculty as will, then there is no way for it to *be* – large, small, heavy, light, free, or unfree. As Dewey says, "what men and women have fought and died for in the name of freedom is multiple and various. But it has never been metaphysical freedom of the will." This was true when Dewey said it in 1922, and it is true today. So change the subject; there is no such thing as free will.[14]

Nevertheless, persons make choices. Typically, they do so with live options before them. If new reasons present themselves, they can change course.[15] If not, they do what they choose or intend. There is a phenomenology to these activities and processes; persons experience themselves choosing, intending, and willing. Ethics sees persons as choosing and thus works the quarry looking for veins of voluntary action involving reasoning, deliberation, and choice. Moral practices of shaping character, of assigning praise and blame, work over the topography of voluntary, involuntary, and nonvoluntary actions, and various admixtures of these. Aristotle and Confucius saw how this worked with zero help from the fiction of "will" – "free" or otherwise.

Understanding Morality

The *genealogy of morals* asks how moral sensibilities, moral values, moral norms, and so on, originate and develop. There is some consensus among naturalistically oriented philosophers that some combination of cultural anthropology, the psychology of learning, and evolutionary biology will play a key role in providing a genealogy of morals. Dewey is helpful: "For practical purposes morals means customs, folkways, established collective habits. This is a commonplace of the anthropologist, though the moral theorist generally suffers from an illusion that his own place and day is, or ought to be, an exception" (Dewey 1922, 55).

In the following passage from *The Descent of Man*, Darwin suggests the general form that an adequate genealogy of morals might take:

> In order that primeval men, or the ape-like progenitors of man, should become social…they must have acquired the same instinctive feelings…They would have felt uneasy when separated from their comrades, for whom they would have felt some degree of love, they would have warned each other of danger, and have given mutual aid in attack or defence. All this implies some degree of sympathy, fidelity, and courage…[T]o the instinct of sympathy…it is primarily due that we habitually bestow both praise and blame on others, whilst we love the former and dread the latter when applied to ourselves; and this instinct no doubt was originally acquired, like all the other social instincts, through natural selection…[W]ith increased experience and reason, man perceives the more remote consequences of his actions, and the self-regarding virtues, such as temperance, chastity, etc., which during earlier times are…utterly disregarded come to be highly esteemed or even held sacred…Ultimately our moral sense or conscience

13 See, for example, Wegner (2002).

14 See Flanagan (2002).

15 An exception is what Dan Dennett (1984) calls "pockets of local fatalism."

> becomes a highly complex sentiment – originating in the social instincts, largely guided by the approbation of our fellow-men, ruled by reason, self-interest, and in later times by deep religious feelings, and confirmed by instruction and habit. (Darwin 1871/2004)[16]

Of course, the genealogical story – both ontogenic and phylogenic – is even more complex than Darwin sketches. Morality evolved and developed in order to coordinate and harmonize the interests (both self- and other-regarding) of humans living in mutually dependent communities. Such communities would need to regulate conflicts of interest, divisions of labor, and hierarchy arrangements, and systems of moral norms would help make such cooperative projects beneficial. On the self-regarding side, morality evolved to shape character and specify worthwhile lives and ideals of behavior toward which to strive. To understand the full story, we will need what we only have pieces of – namely, insights from evolutionary biology, animal ethology, developmental psychology, learning theory, psychiatry, cognitive neuroscience, and cultural anthropology. All these disciplines and research programs are essential to (and thus have a say in) the genealogy.

Normative ethics, on the other hand, is concerned with articulating and defending which virtues, values, norms, and principles will reliably guide favorable character development, intra- and interpersonal well-being, social coordination, and harmony. Normative ethics involves saying and justifying what is right, wrong, good, or bad. Murder and rape are wrong, honesty is the best policy, and so on. Ethical naturalists evaluate their subject matter using standards that are derived from certain human needs, desires, and purposes. Some of these might be thought of as fixed by our natures as social animals; humans need peace, security, friendship, and so on. The specific form of these needs, the ways they are best met, will have a culturally variable component. Some aims or needs are quite culturally specific and defensible. The aims(s) of morality are thus included as part of what humans need and desire.

Naturalistic Epistemology and the Problem of Normativity

In "Epistemology Naturalized," Quine (1969) suggested that epistemology be assimilated to psychology. Many have read Quine's arguments for naturalization as arguments against a normative role for epistemology. Hilary Putnam writes: "The elimination of the normative is attempted mental suicide…Those who raise the slogan '*epistemology naturalized*'… generally *disparage* the traditional enterprises of epistemology" (1993, 229). And Jaegwon Kim writes: "If justification drops out of epistemology, knowledge itself drops out of epistemology. For our concept of knowledge is inseparably tied to that of justification…itself a normative notion" (1993, 224–225).

The alleged problem with epistemology naturalized is this: psychology is not in general concerned with *norms* of rational belief, but with the description and explanation of mental performance and mentally mediated performance and capacities. However, the best way to

16 It is worth marking the fact that when Darwin speaks of "deep religious feelings" being involved in the development of conscience, he points to an anthropological commonplace that we will try to explain. The naturalist may need to accept that humans have dispositions that easily yield religious beliefs and feelings and that these are widely utilized to produce moral motivation. This, of course, is different from saying such beliefs are warranted or true.

approach epistemology naturalized is not to think of epistemology as a "chapter of psychology," where psychology is understood merely descriptively, but rather to think of naturalized epistemology as having two components: a *descriptive-genealogical* component and a *normative* component. Furthermore, not even the descriptive-genealogical component will consist of purely psychological generalizations, for much of the information about actual epistemic practices will come from biology, cognitive neuroscience, sociology, anthropology, and history – from the human sciences broadly construed. More obviously, *normative epistemology* will not be part of psychology, for it involves the gathering together of norms of inference, belief, and knowing that lead to success in ordinary reasoning and in science. And the evolved canons of inductive and deductive logic, statistics, and probability theory most certainly do not describe actual human reasoning practices. These canons (take, e.g., principles governing representative sampling and warnings about affirming the consequent) come from abstracting successful epistemic practices from unsuccessful ones. The database is, as it were, provided by observation of humanity, but the human sciences do not (at least as standardly practiced) involve extraction of the norms. Thus, epistemology naturalized is not epistemology psychologized *simpliciter*.[17] However, since successful practice – both mental and physical – is the standard by which norms are sorted and raised or lowered in epistemic status, pragmatism reigns.[18]

The natural objection here is that all the epistemological work has been done in identifying the cognitive *aims*, and that the relevance of empirical work is just in identifying what *accomplishes* those aims. The two projects seem distinct, with neither one affecting the other. How is this "psychologizing" epistemology to any extent whatsoever? As we noted earlier, a naturalistic approach works back and forth between the normative and the descriptive. Our aims are various and are capable of specification on many levels. Given naturalism's methodology, we can modify our cognitive aims – at least on the more specific levels – with knowledge of how our minds work.[19]

Naturalistic Ethics and the Problem of Normativity

The same worries that Putnam and Kim express over Quine's conception of naturalizing epistemology recapitulate Kant's worries over Hume's approach to naturalizing ethics. John McDowell's criticism of "bald naturalism" in favor of "second nature naturalism" is arguably a way of stating the same concern: namely, that at least some kinds of naturalism are not equipped to explain ethical normativity.[20] In any case, moral psychology, sociology, and

17 Alvin Goldman (1986; 1992) has produced arguably the best work in naturalized epistemology. Goldman never tries to derive normative conclusions from descriptive premises. Furthermore, he continually emphasizes the historical and social dimensions of epistemology in a way that Quine perhaps has not. See also Kornblith (1994).

18 In epistemology, pragmatic evaluation is done relative to our cognitive aims. These, to be sure, are themselves norms, and, as such, are subject to the same sort of requests for rationales and warrant as all other norms.

19 If, to take one example, connectionism is right, then this must have implications for the nature of successful inquiry, and in particular what constitutes successful reasoning.

20 Though our approach is different from McDowell's, we don't think we can be charged with bald naturalism. Actually, it's unclear who he would think is an actual bald naturalist in ethics – unless he is thinking of John Mackie, possibly A.J. Ayer, and perhaps some evolutionary psychologists. We'd be interested in hearing him name names.

anthropology (what Kant called the "empirical side of morals") might tell us what individuals or groups *think* ought to be done, what they *believe* is right or wrong, what they *deem* a good person, and so on. However, all the human scientific facts taken together, including that they are widely and strongly believed, could never *justify* any of these views.

But we should conceive naturalistic ethics in pretty much the same way we conceive naturalized epistemology. Naturalistic ethics will contain a descriptive-genealogical component that will specify certain basic capacities and propensities of *Homo sapiens* – for example, sympathy, empathy, egoism, and so on – relevant to moral life. It will explain how people come to feel, think, and act about moral matters in the way(s) they do. It will explain how and in what ways moral learning, engagement, and response involve the emotions (and which emotions). It will explain what moral disagreement consists in and why it occurs, and it will explain why people sometimes resolve disagreement by recourse to agreements to tolerate one another, without, however, approving of one another's beliefs, actions, practices, and institutions. It will tell us what people are doing (or trying to do) when they make normative judgments. And finally, or as a consequence of all this, it will try to explain what goes on when people try to educate the young, improve the moral climate, propose moral theories, and so on.

Defenders of naturalistic ethics are continually asked to explain how a better picture of moral psychology *can* contribute to our understanding of ethical theory in general and normative ethics in particular. Moral psychology, cognitive science, cultural anthropology, and the other mental and social sciences can tell us perhaps how people *in fact* think and behave. Ethical theories tell us what the aims of ethics are, where to look to ground morality, and so on, while normative ethics tells us how we *ought* to feel, think, and act. It is hard to see, the objectors claim, how such factual or descriptive knowledge can contribute to the project of understanding the aims of ethics, where the sources of moral motivation lie, and how we ought to live.

First, it should be pointed out that *every* great moral philosopher has put forward certain descriptive-genealogical claims, certain theories of philosophical psychology that postulate basic human dispositions which help or hinder morality (e.g., reason, emotion) and the sources of moral motivation. This is a ubiquitous feature of the moral tradition; *everyone* thinks that philosophical psychology (including, e.g., philosophical anthropology) has *implications* for ethics. And though most of these claims suffer from sampling problems and were proposed in a time when the human sciences did not exist to test them, they are almost all testable – indeed, some have been tested (Flanagan 1991). For example, here are four claims familiar from the history of ethics which fit the bill of testable hypotheses relevant to normative ethics: (1) He who knows the good does it; (2) If you (really) have one virtue, you have the rest; (3) Morality breaks down in a roughly linear fashion with breakdowns in the strength and visibility of social constraints; and (4) In a situation of profuse abundance, innate sympathy and benevolence will "receive tenfold increase" and the "cautious, jealous virtue of justice would never once have been dreamed of" (Hume 1748/1751/1975). Presumably, how the descriptive-genealogical claims fare matters to the normative theories and would have mattered to their proponents.[21]

If this much is right, the question arises as to why the contemporary movement to naturalize ethics raises so many hackles. It is true that *philosophical psychology* – the sort that

21 The descriptive-genealogical component will itself be normative in one sense: it will involve descriptions of human actions (etc.) and thus traffic in intentional description. However, it will not be normatively *ethical*.

can be done from an armchair and which is an assemblage of virtually every possible view of mind – is now giving way to *scientific psychology*, which may eliminate some of the classical views of mind on empirical grounds. If this happens, then our ethical theories will be framed by better background theories about our natures. What could be wrong with this?

Hume's Objection

The standard view, again, is that nothing normative *follows* from any set of descriptive-genealogical generalizations. David Hume is supposed to be the father of this line of objection. Yet Hume (along with Aristotle) is often thought to be a father of ethical naturalism. Could it be that Hume was objecting to his own enterprise? In fact, Hume's "objection" is limited to one paragraph in his *Treatise of Human Nature*:

> In every system of morality, which I have hitherto met with…the author proceeds for some time in the ordinary ways of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, is and is not, I meet with no proposition that is not connected with ought or ought not…[A]s this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that shou'd be observ'd and explain'd…how this new relation can be a deduction from others, which are entirely different from it. (Hume 1739/1978)

Hume is making a rather simple point here: one cannot draw normative conclusions from non-normative premises alone. But Hume found the slide from statements of mere fact to statements of value to be characteristic of "vulgar systems of morality." For him, morality was inexplicable without mentioning moral sentiments or passions.[22]

Let us return, then, to the objection in its current form: namely, that nothing normative follows from any of the empirical information we might gather from the natural, social, and human sciences. Perhaps the fear is that if the background theory is scientific, this makes ethics a science, or that if the background theory is a science, we can suddenly violate the laws of logic and derive "ought" from "is." However, no one has suggested these things! No important moral philosopher, naturalist or non-naturalist, has ever thought that merely gathering together all relevant descriptive truths would yield a full normative ethical theory. Morals are radically underdetermined by the merely descriptive, but so too, of course, are science and normative epistemology. All three are domains of inquiry in which ampliative generalizations and underdetermined norms abound.

The smart naturalist makes no claims to establish demonstratively moral norms. Instead, he or she points to certain practices, values, virtues, and principles as reasonable based on inductive and abductive reasoning (more on this further on). Indeed, anyone who thinks that Hume thought that the fallacy of claiming to move demonstratively from is to ought revealed that normative ethics was a nonstarter hasn't read Hume. After the famous passages in the *Treatise* about is–ought, Hume proceeds for several hundred pages to do normative moral philosophy. He simply never claims to *demonstrate* anything. Why should he? Demonstration, Aristotle taught us long ago, is for the mathematical sciences, not for ethics.

22 See Sutherland and Hughes (2000).

Moore's Fallacy

Regarding the challenges to naturalism based on open-question arguments, the ethical naturalist has all the resources to effectively meet the challenges. Ethics naturalized need not be reductive, so there is no need to define "the good" in some unitary way such that one can ask the allegedly devastating question: "But is that which is said to be 'good,' good?"

Indeed, the force of open-question arguments fizzled with discoveries about failures of synonymy across the board – with discoveries about the lack of reductive definitions for most interesting terms. Suppose "good" is taken to be a term and we think, as Moore does, that it should have a definition. If Moore thinks (as it seems he does) that a definition ought to supply necessary and sufficient conditions of application, then he is correct that if such a definition were available it would not be "open" to "questioning." However, except for some technical terms (e.g., "even numbers" and "odd numbers") and certain scientific terms, most others do not have necessary and sufficient conditions of application. What we call "dictionary definitions" are a mix of current usage patterns and functional characterizations. This makes sense, given that most terms in natural language have some sort of prototype/exemplar/stereotype structure.[23] It is not surprising that G.E. Moore couldn't find a definition of "good." Failure to find a definition of "good" would no more prove that it names a non-natural property than the same failure to find definitions for "fuzzy" or "chair" would prove that fuzziness is a non-natural property or that chairs are non-natural objects.

Moreover, "good" is not a singular term in our language, including the moral sense of "good." Instead, it is a theoretical term in the following sense: we call different things morally good for different reasons. Moral virtues, for example, are morally good for reasons r_1, $r_2 \ldots r_n$, and so on. It takes a complex moral conception to help fix the multiple meanings of "morally good." "Morally good" likely names a heterogeneous set.

Relativism and Nihilism

This leads to some final alleged obstacles to naturalism – namely, that it typically leads to relativism, that it is deflationary and/or morally naive, and that it makes normativity a matter of power: either the power of benign but less than enlightened socialization forces or the power of those in charge of the normative order (possibly fascists, Nazis, or moral dunces).

How does naturalistic ethics avoid extreme relativism or – even worse – nihilism? The answer is simple: *the ends of creatures constrain what is good for them*. The relativist is attuned to relations that matter; not all kinds of food, clothing, and shelter suit us members of the species *Homo sapiens*. Nor do all interpersonal and intrapersonal practices suit us. Thus, there are substantial constraints on what might count as an adequate morality stemming from intrapersonal and interpersonal factors.[24] We are social animals with certain innate capacities and interests. Though the kinds of play, work, recreation, knowledge, communication, and friendship we seek have much to do with local socialization, the general fact that we like to play, work, recreate, know, communicate, and befriend seems to be part, as we say, of human nature.

23 See Casebeer (2003) for a similar argument against Moore. Prinz (2002) gives an edifying discussion of concepts and their prototype-like structure.

24 A more detailed account can be found in David Wong's (2006b) *Natural Moralities*.

The distinctively normative component of naturalistic ethics should explain *why* some norms (including norms governing choosing norms), values, and virtues are good or are better than others. One common rationale for favoring a norm or set of norms is that it is suited to modify, suppress, transform, or amplify some characteristic or capacity belonging to our nature – either our animal nature or our nature as socially situated beings. Consider some of the core moral beliefs likely to be found across various cultures: beliefs concerning the permissibility of killing, rights to property and resources, and the need for norms of reciprocity. These beliefs might vary from culture to culture, but they all serve to regulate and promote human social life. Even prior to the powerful (natural) effects of culture, we prefer different things when it comes to shelter, play, communication, and friendship than do beavers, otters, dolphins, birds, orangutans, and bonobos. Morality cannot seek to instantiate behavior that no human beings have a propensity to seek. This much constrains extreme relativism.

This seems to reduce morality to a system of *hypothetical imperatives* that hinge on our wanting to secure certain aims: "*If* you want to secure social cooperation, *then* you ought to…" It is true that naturalists cannot allow for *categorical imperatives* if they are conceived of as independent of human interests and values, or categorical imperatives that are binding on all rational beings, wherever they might be. Yet, while the aims of naturalistic ethics are *internal* to the motivational systems of the species *Homo sapiens*, they are *external* to any particular individual member of that species.[25] This follows from the view that there are a limited number of goods that human beings seek, given their nature and potentialities, and these goods (or aims) limit what can be placed as antecedents to the hypothetical conditionals. In referring to these facts in moral discourse, one is not simply pointing to preexisting propensities in any given individual, but rather is referring to basic and fundamental reasons stemming from human nature that might help shape and channel the particular propensities of any given individual. In this sense, they do have some "categorical" force.[26] Pluralistic relativism articulates and advances a theory about the constraints on "morally adequate" plural ways of life that aim at the set (or some subset) of the goods that constitute morality, broadly construed.

Nihilism is also not a problem. Humans seek value; we aspire to goods, to things that matter and interest us. Now, nihilism can be a problem for individuals when their *motivating* reasons (discussed earlier) are exposed as not good *grounding* or *justifying* reasons. (The loss of faith in parental wisdom and authority during adolescence is one example.) Nihilism is also a familiar problem for theists who lose their faith, and for very depressed humans for whom things have stopped mattering. However, nihilism is not a special problem for naturalists. Animals like surviving; reflective animals like living well. Over world-historical time, reflective animals develop goals for living: welfare, happiness, love, friendship, respect, personal and interpersonal flourishing. These are not an altogether happy and consistent family of values. Still, even if there are incompatibilities involved among the ends we as animals – socialized animals – seek, the fact remains that there *are* ends we seek, and nihilism is not normally an issue; it is not usually a "live option." Nihilism is the view that *nothing* matters. Things do *matter* for us: certain things matter because of our membership in a certain biological species, and certain things in virtue of how we have evolved as social beings with a history. That is the way it is.

25 For more on this view, see Wong (2006a,b).

26 For another take on the problems of categoricity and nihilism from a naturalistic perspective, see Railton (1986). Railton believes moral imperatives apply even to those who have no reason to follow them, because rationality is not a precondition of moral obligation. This represents one way of introducing categoricity that is different from our own approach. Railton uses a similar tactic to evade nihilism.

Ethics Naturalized: Pluralism and Human Ecology

We close with Dewey's insight that "Moral science is not something with a separate province. It is physical, biological, and historic knowledge placed in a humane context where it will illuminate and guide the activities of men" (1922, 204–205). What is relevant to ethical reflection is everything we know, everything we can bring to ethical conversation that merits attention. To put a pragmatist spin on the point, we can say that moral knowledge obeys the canons of inductive and deductive logic, statistics, and probability theory in producing warranted beliefs about which traits are virtues, which are vices, and what values, actions, norms, and principles reliably produce social coordination and human flourishing. The normative component involves the imaginative deployment of information from *any* source useful to self/social examination, forming new or improved norms and values, improving moral educational practices, training moral sensibilities, and so on. These sources include psychology, cognitive science, and all the human sciences (especially history and anthropology), as well as literature, the arts (for the arts are ways of knowing and ways of expressing insights about our nature and about matters of value and worth),[27] and ordinary everyday observations about how individuals, groups, communities, nation states, and the community of persons or sentient beings are faring. The aims relevant to these sorts of pragmatic evaluation are various and are capable of specification on various levels. First-order, second-order, third-order, and possibly higher-order level evaluation of norms are things natural human minds can do. We can tinker with these aims, systematizing them where this proves useful, addressing severe conflicts (e.g., between universal and particular duties), and thinking of possible ways of addressing them (e.g., through revising our norms). The pragmatist is committed to the requirement that normative judgments get filled out in conversation and debate, that his or her background criteria are open to criticism and reformulation, and that terms like "what works" and "what conduces to flourishing" are superordinate. Specificity is gained in more fine-grained discussion of particular issues. Even if there is no such thing as "transcendent rationality," no ultimate or non-question-begging way of establishing one's viewpoint over another, there are perfectly reasonable ways of analyzing problems and proposing solutions.

If ethics is like any science or is part of any science, it is part of *human ecology*, concerned with saying what contributes to the well-being of humans, human groups, and human individuals in particular natural and social environments. What is good depends a great deal on what is good for a particular community, but when that community interacts with other communities, these too get a say. Furthermore, what might *seem* like a good practice or ideal can, when all the information from history, anthropology, psychology, philosophy, and literature is brought in, turn out not to have been such a good idea after all. If ethics is part human ecology, then the norms governing the evaluation of practices and ideals will have to be as broad as possible. To judge ideals, it will not do simply to look and see whether healthy persons and healthy communities are subserved by them in the here and now; it must also be the case that this "health" is bought without incorporating practices – slavery, racism, sexism, and the like – which we know can go unnoticed for some time but that can

27 Richard Rorty (1991) convincingly suggests that the formulation of *general moral principles* has proven less useful to the development of liberal institutions than has the gradual *expansion of the imagination*, for example through the writings of individuals such as Friedrich Engels, Harriet Taylor and J.S. Mill, Harriet Beecher Stowe, Bronislaw Malinowski, Martin Luther King, Jr., Alexis de Tocqueville, and Catherine MacKinnon.

keep persons from flourishing and eventually poison human relations, if not in the present, at least in nearby generations.

The aims of morality are heterogeneous, not always individually or collectively or at all times optimifically satisfiable. One aim of ethics is to try to make the best of this fact in particular ecological niches, particular historical communities with their own sets of aims, practices, and so forth. Thinking of normative ethical knowledge as something to be gleaned from thinking about human good relative to particular ecological niches will make it easier for us to see that there are forces of many kinds, operating at many levels, as humans seek their good; that individual human good can compete with the good of human groups and of nonhuman systems; and finally, that only some ethical knowledge is global: most is local, and appropriately so. It might also make it seem less compelling to find ethical agreement where none is needed.

The localized and contingent nature of many of the values we hold dear is no reason for not cherishing them, no reason to deny them a constitutive role in providing meaning. There are some things to be said for contingency (besides the fact that consciousness of it can possibly undermine confidence, self-respect, and their suite). Recognition of contingency has the advantage of being historically, sociologically, anthropologically, and psychologically realistic. Realism is a form of authenticity, and there is much to be said in authenticity's favor. Furthermore, recognition of contingency can engender respect for human diversity, which engenders tolerant attitudes. This has generally positive moral and political consequences. And this is all consistent with deploying our critical capacities in judging the quality and worth of alternative ways of being. Attunement to contingency, plural values, and the vast array of possible human lives and personalities opens the way for use of important and underutilized human capacities: capacities for critical reflection, for seeking deep understanding of alternative ways of being and living, and for deploying our agentic capacities to modify ourselves by engaging in identity experimentation and meaning location within the vast space of possibilities that have been and are being tried by our fellows. There are many things to be said in favor of emphasizing "consciousness of contingency."

The pluralistic relativist, the pragmatic human ecologist, has the right attitude: right for a world in which profitable communication and politics demand respect and tolerance, but in which no one expects a respectful, tolerant person or polity to lose the capacity to identify and resist evil where it exists, and right in terms of the development of our capacities of sympathetic understanding, acuity in judgment, and self-modification – and, on occasion, radical transformation.

Acknowledgments

Our thanks to Walter Sinnott-Armstrong for many helpful comments on drafts of this chapter.

References

Audi, R. (ed.) (1999). *Cambridge Dictionary of Philosophy*, 2nd edn. Cambridge: Cambridge University Press.

Bouwsma, O.K. (1948). Naturalism. *Journal of Philosophy* 48: 12–21.

Casebeer, W.D. (2003). *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*. Cambridge, MA: MIT Press.

Chisholm, R. (1966). Freedom and Action. In *Freedom and Determinism*, edited by K. Lehrer, pp. 11–44. New York: Random House.

Craig, W.L. and Sinnott-Armstrong, W. (2003). *God? A Debate between a Christian and an Atheist*. New York: Oxford University Press.

Darwin, C. (1871/2004). *The Descent of Man, and Selection in Relation to Sex*. London: Penguin.

Dennett, D.C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.

Descartes, R. (1649/1968). *Passions of the Soul*. In *The Philosophical Works of Descartes*, eds. E. Haldane and G. Ross. Cambridge: Cambridge University Press.

Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Modern Library.

Flanagan, O. (1991). *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge, MA: Harvard University Press.

Flanagan, O. (1995). Ethics Naturalized: Ethics as Human Ecology." In *Mind and Morals*, edited by Larry May, Marilyn Friedman, and Andy Clark. Cambridge, MA: MIT Press.

Flanagan, O. (2002). *The Problem of the Soul*. New York: Basic Books.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, A. (1992). *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.

Greene, B. (2000). *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: Vintage Books.

Honderich, T. (ed.) (1995). *The Oxford Companion to Philosophy*. New York: Oxford University Press.

Hume, D. (1739/1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hume, D. (1748/1751/1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press.

Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.

Kant, I. (1785/1964). *Groundwork of the Metaphysic of Morals*, translated and analyzed by H.J. Paton. New York: Harper & Row.

Kim, J. (1993). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kitcher, P. (1992). The Naturalists Return. *Philosophical Review* 101.1: 53–114.

Kornblith, H. (1994). *Naturalizing Epistemology*, 2nd edn. Cambridge, MA: MIT Press.

MacIntyre, A. (1988). *Whose Justice? Which Rationality?* Notre Dame: University of Notre Dame Press.

MacIntyre, A. (1991). *Three Rival Versions of Moral Inquiry*. Notre Dame: University of Notre Dame Press.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin Books.

McDowell, J. (1996). *Mind and World*. Cambridge, MA: Harvard University Press.

Pigden, C. (1991). Naturalism. In *A Companion to Ethics*, edited by Peter Singer, pp. 421–431. Oxford: Blackwell.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

Putnam, H. (1993). Why Reason Can't Be Naturalized. In *Philosophical Papers, Vol. III: Realism and Reason*. Cambridge: Cambridge University Press.

Putnam, H. (2004). *Ethics Without Ontology*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Railton, P. (1986). Moral Realism. *Philosophical Review* 95.2: 163–207.

Rorty, R. (1991). On Ethnocentrism. In *Philosophical Papers*, vol. 1. Cambridge: Cambridge University Press.

Rosenberg, A. (1996). A Field Guide to Recent Species of Naturalism. *British Journal of the Philosophy of Science* 47: 1–29.

Stroud, B. (1996). The Charm of Naturalism. *Proceedings and Addresses of the American Philosophical Society* 70: 43–55.

Sutherland, K. and Hughes, J. (2000). Is Darwin Right? *Journal of Consciousness Studies* 7(7): 63–86.

Taylor, C. (1989). *Sources of the Self*. Cambridge: Cambridge University Press.

Wegner, D.M. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wong, D.B. (1984). *Moral Relativity*. Berkeley: University of California Press.

Wong, D.B. (1996). Pluralistic Relativism. *Midwest Studies in Philosophy* 20: 378–400.

Wong, D.B. (2006a). Moral Reasons: Internal and External. *Philosophy and Phenomenological Research* 72(3): 436–558.

Wong, D.B. (2006b). *Natural Moralities*. Cambridge: Cambridge University Press.

3

Naturalism in the Continental Tradition

KEITH ANSELL PEARSON AND JOHN PROTEVI

Though the term "naturalism" is used more by analytic philosophers than by continental philosophers, the issues denoted by it are of great concern to continental philosophers. In this chapter, we will use the following rough-and-ready terminology. We begin by distinguishing methodological and ontological naturalism. Methodological naturalism (MN) varies in strength. Weak MN asserts a compatibility between the goals and methods of philosophy and natural science, so that they can and should inform each other, even though philosophy maintains its specificity and independence from science; strong MN asserts a continuity between philosophy and natural science, so that philosophy has no real independence (i.e., there are no specifically philosophical questions in continuity MN); superstrong MN, or "scientism," asserts that only natural science produces meaningful statements, credence-worthy beliefs, and so on. Ontological naturalism (ON) also varies in strength. Weak ON merely denies the existence and explanatory power of supernatural entities, but allows entities and states emergent from and irreducible to the physical (such as organisms), to the biological (such as social institutions), and to the neurological (such as mental states); strong ON is nonreductive physicalism, allowing vital, social, or mental properties of physical events, but not vital, social, or mental states, and thereby asserting a property dualism linked to a substance monism; superstrong ON, or "(reductive) physicalism," asserts that only physical entities exist (thereby asserting the identity of vital, social, and mental states with physical states and, in its eliminative posture, the ultimate dispensability of the notion of vital, social, or mental properties).

Though we will discuss MN and ON, we will not divide this chapter along those lines; rather, we will discuss antinaturalism in phenomenology and critical social theory as bookends to a discussion of affirmative naturalism. To forecast the paper: we will treat phenomenology and critical social theory somewhat briefly in order to have room to discuss affirmative naturalism. It seems defensible to us, in a volume on naturalism, to stress the positive uptake of naturalism in continental philosophy. We will begin by treating the antinaturalism of Edmund Husserl's phenomenology, followed by the (weak MN)

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

challenge of "naturalizing phenomenology." As a transitional figure, we will treat Hans Jonas and the weak ON status he allows organismic life. In affirmative naturalism, we will treat Friedrich Nietzsche, Henri Bergson, and Gilles Deleuze. We will conclude by considering the antinaturalism of continental philosophy positions in critical race theory, gender theory, and disability studies.

Phenomenological Antinaturalism

Strong (continuity) and superstrong (scientism) MN are little represented in continental philosophy, though many affirmative naturalists accept a version of weak (compatibility) MN, in which they discuss and indeed incorporate scientific findings into their philosophy. With the exception of the weak MN position of Maurice Merleau-Ponty (Zahavi 2010), phenomenologists are actively hostile to all versions of MN and ON; the prime example of this hostility is Edmund Husserl, the founder of phenomenology as a philosophical school. Phenomenology in this sense is not simply the description of first-person experience (which is how the term is used in AP), but instead the methodologically self-aware analysis of the conditions of meaningful appearance of the world.

Husserl's Antinaturalism

To understand Husserl's antinaturalism, we must see him in the context of post-Kantianism. There were three main responses to Kant's critical project: German idealism; naturalization via scientific psychology; and neo-Kantianism. Opposition to German idealism propelled the naturalizers and the neo-Kantians. Among the naturalizing psychologists, Helmholtz was the most celebrated; for him and the others, the Kantian transcendental conditions of knowledge (space and time as the *a priori* forms of intuition and the categories as the *a priori* concepts of the understanding) are naturalized by being treated physiologically, so that sensations are signs or symbols that are translated by the brain into experience (Friedman 2000; Luft and Capeillères 2010; Hatfield 2012). F.A. Lange, author of the influential *History of Materialism* (1866/1974), adds an evolutionary perspective: the subjective conditions of possibility of experience are species-specific adaptations (see Stack 1991 and Cox 1999 for considerations of the influence of Lange on Nietzsche).

Neo-Kantianism was the dominant school of German philosophy at the time of Husserl's training and early publications (Makkreel and Luft 2010). Neo-Kantians shared with the naturalizers the rejection of German idealism; they were, however, antinaturalist insofar as they were antipsychologist, as, for them, ideal, universally valid, scientific facts cannot depend upon empirical psychological processes. Hermann Cohen of the Marburg School of Neo-Kantianism went so far as to claim that the *a priori* laws of human knowledge as conditions of objectivity are ultimately the laws of physics (and the faculties of cognition are the methods of mathematics and physics); for Cohen, an object of experience is not the star a person sees but the orbit an astronomer calculates (Luft and Capeillères 2010).

Following Frege's criticism of his early work, Husserl was staunchly antipsychologistic in his breakthrough text, the *Logical Investigations* (Husserl 1900–01/1973; Mohanty 1984; Kusch 2014), but in his mature work he affirmed transcendental subjectivity as world-constituting. The key Husserlian term, then, is "transcendental," which for him retains a Kantian sense of providing the conditions for experience without itself being empirical.

Transcendental phenomenology is nonetheless more closely related to the empirical than was the case for Kant, given that phenomenology begins with describing and analyzing the concrete ego's experience, rather than sketching the sheer conditions for any rational experience (Zahavi 2010; see Lawlor 2002 for Derrida's radicalization of the "zig-zag" relation between transcendental and empirical). Whatever the complications his descendants developed, Kant himself was anti-MN in distinguishing transcendental philosophy from natural science; transcendental philosophy is a form of knowledge irreducible to, and providing conditions of possibility for, natural science. For Husserl, then, transcendental phenomenology retains the connection to subjectivity, but it must be rescued from empirical psychology.

For the Husserl of the 1910s, we need to bracket the "natural attitude" –roughly speaking, a realist position taking for granted the existence of worldly objects – that underlies both everyday life and natural science. So Husserl is anti-MN in that phenomenology is not a natural science, nor can its results ever be arrived at via natural science. However, phenomenology is itself a science of consciousness grounding the natural sciences; it is thus a more fundamental science than the natural sciences (Husserl 1910/1965; Moran 2008). In classical Husserlian transcendental phenomenology, we cannot see the subject as just another object, as we would in the natural attitude. Thus, we need to distinguish the transcendental from the empirical subject; with regard to its anti-ON position, phenomenology does not see transcendental consciousness as a thing in the world, but it is also not supernatural in any mysterious, divine, spooky, or ghostly sense. Rather, it is the condition of meaning of the world as it appears; it is accessed by a change in perspective, not by a flight into the heavens (Zahavi 2004).

In the 1930s, the time of the *Crisis* (1954/1970), Husserl's anti-MN position crystallizes around what he calls "vague essences." Vague essences (e.g., "roundness") are midway between round things ("wheels") and geometrical objects ("circle"). They are oriented toward scientific objectivity, which they reach by a passage to the limit, but they are not proper objects of geometry; they could never be quantified as would be required, on Husserl's understanding, for their naturalization as scientific objects (anti-MN). They are encountered in the life world, a purification of which is needed to reach scientific objects; and the life world can be explored only by phenomenology (Moran 2008; Lawlor 2010). So natural science – in this case, psychology – could never exhaust the realm of objects revealed by phenomenology (anti-ON).

The Naturalizing Phenomenology *Challenge*

As is clear from its title, the essay collection *Naturalizing Phenomenology* challenged phenomenology's antinaturalism (Petitot et al. 1999; see also Varela 1996 for the term "neurophenomenology"). The volume proposes that the state of mathematics at the time drove Husserl to his antinaturalism:

> What Husserl called "inexact morphological essences," essences foreign to fundamental classical physics, are indeed amenable to a physical account, provided that we rely upon the qualitative macrophysics of complex systems (and no longer upon the microphysics of elementary systems). (Petitot et al. 1999, 55)

It is true, the editors of the *Naturalizing Phenomenology* volume concede, that one can never get precise quantitative predictions of microstate models of massively complex systems like the brain or the brain–body–world system. But with dynamic systems models, they claim, one can get qualitative predictions of mesostate models.

Reactions to the volume by those sympathetic to phenomenology were mixed, with some (e.g., Zahavi 2004) taking issue with the details of the volume's approach but allowing for some ways in which the project could be useful, and others (e.g., Lawlor 2010) rejecting the project entirely. Both Zahavi and Lawlor emphasize Husserl's principled philosophical antagonism to naturalism, which they insist was not simply a contingent opposition based on the empirical state of mathematics of his time. Nonetheless, Zahavi mentions "a way out," noting that Husserl had thematized "phenomenological psychology" as a preliminary approach to transcendental phenomenology. According to Zahavi, "Phenomenological psychology is…a form of descriptive, eidetic, and intentional psychology which takes the first-person perspective seriously, but which…remains within the natural attitude. [It] might be described as a local regional-ontological investigation, which investigates consciousness for its own sake" (2004, 339). Thus, phenomenological psychology, while it is not as fully philosophical as transcendental phenomenology, can nevertheless aid cognitive science by providing it exact descriptions of the target phenomena it tries to explain. As Overgaard puts it:

> Contemporary consciousness research seems to be in the strange situation of not having very sophisticated descriptions of its object of research. Terms like "conscious state" or "experience" are quite imprecise and have several different meanings. For example, "being conscious" could mean anything from "knowing" or "being aware of," to "having a phenomenal experience," to "being awake." For that reason alone, it seems reasonable to turn to phenomenology, which, since the writings of Husserl, has tried to give precise and coherent descriptions of events and structures in consciousness. (2004, 365)

Jonas

Let us consider Hans Jonas as a transition from the antinaturalism of the phenomenologists. Working in the phenomenological tradition, he nonetheless proposes a phenomenology of the organism that is in the neighborhood of weak ON; Jonas wants to specify a biological yet nonreductive physicalist and hence emergent status for organismic life.

In his essay "Is God a Mathematician? The Meaning of Metabolism", Jonas (2003, 64–92) writes that modern mathematical physics gives us time as a series of instants, such that the physical states of a process are externalized, one to the other: "each of them determined anew by the component factors operative at that very instant" (2003, 68). Such fragmentation means that analysis meets no resistance; in other terms, there is no wholeness, only an aggregation of moments, and so ontological emergence is denied (2003, 69).

Testing the reduction of biology to physics, Jonas proposes the wave as the physicist's model of complex physical form, a form that is wholly reducible to an aggregate. The wave, as an "integrated event-structure" has no ontologically emergent status, and what is true of the wave must be true of the organism as object of divine intellection. Without need of the "fusing summation of sense," for God, "the life process will then present itself as a series, or a web of many series, of consecutive events concerning these single, persisting units of general substance" (2003, 77). For Jonas, however, such a reductive account misses the ontological emergence that makes of life an "ontological surprise," and makes of the organism a system, a "unity of a manifold." The organism is "whole" as "self-integrating in active performance," an "active self-integration of life" (2003, 79). The "functional identity" of an organism relative to the materials it metabolizes is constituted "in a dialectical relation of *needful freedom* to matter" (2003, 80; emphasis in original). Both elements, need and

freedom, constitute the "transcendence" of life, and this transcendence constitutes a living present, a metabolically founded transcendental aesthetic or *a priori* form of organic time: "self-concern, actuated by want, throws open…a horizon of time…the imminence of that future into which organic continuity is each moment about to extend by the satisfaction of that moment's want" (2003, 85).

Jonas then draws the consequences for the question of the adequacy of purely mathematical physics for the phenomenon of life: "with respect to the organic sphere, the external linear time-pattern of antecendent and sequent, involving the causal dominance of the past, is inadequate." With life on the scene, though, "the extensive order of past and future is intensively reversed," so that the determination of "mere externality" by the past has to be supplemented by the recognition that "life is essentially also what is going to be and just becoming" (2003, 86).

Jonas is anti-strong or superstrong ON in resisting the reduction of life to physics (his position is not physicalism), but he embraces a weak ON in that his discussion of the organism revealed in a phenomenology of life does not rely upon a supernatural "life force," or even a substantive vitalism of the Driesch entelechy school (Wolfe 2010).

Affirmative Naturalism

A weak ON, one that denies the existence of supernatural entities but affirms that of emergent entities and states, is well represented in continental philosophy. Among the lines of thought here would be the Spinozist/Nietzschean/Deleuzean one. Though these thinkers are monists, Spinoza is not strictly speaking a reductive materialist, but rather a property dualist who also allows composite or emergent entities ("corporate" bodies). While Nietzsche and Deleuze are materialists, they are not physicalists or mechanists; in his terms, Deleuze is a "machinic" materialist. We will emphasize not just the metaphysics here but also the ethical component. This has little to do with "ethical naturalism" as it is understood in AP; rather, it is an ethics of joy, in the Spinozist tradition.

Nietzsche and Bergson

Recent works explore Nietzsche as a naturalist (Cox 1999; Leiter 2002; Moore 2002; Hatab 2004; Richardson 2004). Nietzsche's commitment to weak ON is clear: "*My purpose*: to demonstrate the absolute homogeneity of all events" (Nietzsche 1906/1968, #272; italics in original). However, Nietzsche's brand of naturalism does not mean he is a mechanist. For Nietzsche, mechanism leaves notions of "reason" and "purpose" out of the picture as far as possible, showing that, given sufficient time, anything can evolve out of anything else; we can, for example, account for life in terms of pressure and stress. But then he notes that, from this position, we are unable to "explain" pressure and stress themselves, and we cannot get rid of "action at a distance" (1906/1968, #618). In *Beyond Good and Evil*, he appeals to a "conscience of method" to justify his claim that the world seen from inside is "will to power" (Nietzsche 1886/1966, #36); in *Will to Power*, he writes: "…one is obliged to understand all motion, all 'appearances,' all 'laws,' only as symptoms of an inner event and to employ man as an analogy to this end" (1886/1968, #619).

In the case of "life," Nietzsche notes that mere variations of power could not feel themselves to be such, and thus, "there must be present something that wants to grow and

interprets the value of whatever else wants to grow" (1886/1968, #643). Though this admits a limited teleology into our understanding – one that is necessary to our understanding of life in terms of growth and expansion – Nietzsche advises us to beware of "*superfluous* teleological principles," such as positing the instinct of preservation as the cardinal drive (he holds instead that a living thing desires above all to discharge its force) (1886/1968, #650). On this model, then, mechanism and matter are to be construed as expressions of lower stages of life, as "the most despiritualized form of affect (of 'will to power')" (1886/1968, #712).

Nietzsche thinks that this dynamic interpretation of the world, with its denial of empty space and little clumps of atoms, will come to dominate physics; it is in this context that his interest in the example of music needs to be appreciated. In a note entitled "Against the Physical Atom," he writes:

> The calculability of the world, the expressibility of all events in formulas –is this really "comprehension"? How much of a piece of music has been understood when that which in it is calculable and can be reduced to formulas has been reckoned up? And "constant causes," things, substances, something "unconditioned" *invented* – what has one achieved? (1886/1968, #624; see also 1887/1974, #373)

Nietzsche's concern, then, is not simply phenomenological in the sense of what something feels like to me as subject of experience, but an ontological one about the need for a dynamic principle to account for life. Nietzsche's commitment to monism means that he does not need to operate with a distinction between MN and ON. Instead, his focus is on degrees and kinds of difference among material and spiritual forms of life (e.g., differences of complexity in the organization of living systems).

Nietzsche has often been read as a naturalist; Bergson's case is more complex. He is frequently seen as part of the French "spiritualist" tradition, which emerged as an attack on the French positivism of Comte, Taine, and Renan. But the approach he adopts in *The Two Sources of Morality and Religion* is eminently naturalistic, looking at morality and religion on the basis of the "closed" and "open" tendencies of social life (Bergson 1935/1977; see also Lefebvre and White 2012).

As already stated, we will discuss the ethics, as well as the metaphysics, of our affirmative naturalists. To that end, a comparison of Nietzsche and Bergson's naturalist ethics is in order. We begin with the criticisms they make of rationalist and intellectualist approaches. One could argue that the primary category for both, and what links them together, is life. Nietzsche accords ontological and ethical primacy to the will-to-power (which, for him, is what life is), and Bergson does the same with his notion of the *élan vital* (vital impetus). The naturalistic statuses of the will-to-power and *élan vital* are doubtless controversial, but we can at least note Bergson's refusal to assimilate his notion to Driesch's entelechy or individualized life force (Bergson 1911/1998, 42–43; see also Ansell Pearson 2005 for a possible influence of Bergson on Canguilhem); on the possibility of articulating at least some deflationary senses of will-to-power in Darwinian terms, see Ansell Pearson (1997) and Richardson (2004).

There are important influences on both Nietzsche and Bergson from within the tradition of philosophical naturalism: Spinoza and Hume, in particular. In the summer of 1881, on the eve of his experience of eternal recurrence, Nietzsche wrote to a friend that he had discovered he had a precursor, namely, Spinoza, and enumerated the points on which he felt

this affinity: the denial of free will as absence of conditions, the denial of evil and a moral world order, and the commitment to making of knowledge the most powerful passion (cited at Deleuze 1970/1988, 129). What both Bergson and Nietzsche admire in Spinoza is the naturalism of conceiving the human as part of nature, leading to a rejection of transcendence and an affirmation of immanence. However, neither is a straightforward Spinozist, and both regard him as too much of a Stoic, getting worryingly attached to philosophical self-absorption.

Hume is an important influence on both Nietzsche and Bergson, insofar as Hume's naturalism held that we should study the human mind in exactly the same way we study the forces of physical objects (laws of attraction and repulsion, laws of association, etc.; Hume 1740/1978, 645). In *On the Genealogy of Morals*, Nietzsche praises "English psychologists" (in which group he no doubt includes Hume) for promoting, through free inquiry, a set of "ungodly, immoral, and unchristian truths," notably the idea that the workings of the human mind can be explained through such "base" phenomena as the inertia of habit and the random, mechanical coupling or association of ideas (Nietzsche 1887/1967, 1.1). This works against our intellectual pride: the intellect wants to believe that our minds have more lofty sources and superior operations. But then, in *Beyond Good and Evil*, he also criticizes "English philosophy" for its limited "plebeian" ambitions and argues that the likes of Hume and Locke represent a devaluation of the concept of the philosopher (Nietzsche 1886/1966, #252). Bergson's Humean inspiration is no doubt the view that the human is not primarily a rational animal; reason is always subordinate in its operations to feelings and beliefs, which have their source in our nature and are not derived from reason, such that reason is not an independent faculty.

Deleuze

It is customary, though not necessarily all that illuminating, to classify Deleuze as a poststructuralist. In a nod to convention, we will adopt that phrase, thus locating Deleuze after the structuralist challenge to phenomenology. Rather than locating sense in transcendental subjectivity (or in one of its existentialist rivals, such as Heidegger's Dasein or Merleau-Ponty's lived body), the structuralists were, in the words of Paul Ricouer, a sort of "Kantianism without the transcendental subject"; that is, a search for formal and oppositional structures of intelligibility located in cultural systems rather than in a subject. But the structuralist move paid the price of reducing history to an exterior realm that would at most activate switches embedded in the structures; the poststructuralists insist on the historical formation of the unconscious social structures underlying the production of sense (and of desire).

Deleuze's naturalism is evident in his early essay on Lucretius (in Deleuze 1969/1990), in which he praises Lucretius' pluralistic naturalism: an understanding of nature as a distributive rather than a collective power. As Patrick Hayden notes, nature so understood "is that which produces the diverse, yet it does not totalize the diverse into the transcendent One, Whole, or Being to which Platonic anti-naturalism aspires" (1998, 107). For Deleuze, Lucretius thinks of nature in terms of multiplicity, as a nontotalizable sum of diverse individuals, species, and environments. In Lucretian naturalism, our actions are to be guided not by adherence to supernatural myths and illusions, but rather by the affirmation of the positive power of an immanent and multiple nature, and by the joy resulting from the diversity of its elements.

In his major work on Spinoza, Deleuze speaks of a "new naturalism"; in one instance he also recognizes a "new materialism" (1968/1992). The new materialism is, for Deleuze, first and foremost a philosophy of immanence; here he sees the attempt to recognize the positivity of nature and to grant the human being the reflective capacities necessary to penetrate these depths (1968/1992, 322). Deleuze reads Spinoza as proposing an "expressive" nature, a nature of causal explication, and argues for an immanence of expression in what expresses itself (substance and modes). Deleuze acknowledges that Spinoza's immanence "insinuates itself among the transcendent concepts of emanative or creationist theology" (1968/1992, 232). Spinozist substantial monism is radically immanent; there is no transcendence of the One beyond or above Being or transcendence of a Being above its creation. Deleuze thus speaks of Spinoza's "realization of the naturalist program" that has both mechanist and dynamic aspects (1968/1992, 229).

In fact, Deleuze locates the "new naturalism" in both Leibniz and Spinoza, but clearly prefers the latter. Deleuze holds Leibniz's finalism to be an inverted mechanism in which, though there is an expressive nature, this nature is given by God in the form of a preestablished harmony. Things are very different in Spinoza, in whom we find a pure immanent causality, so that, on this conception of nature, finality is excluded. This means that there is no given moral harmony, no metaphysics of essences, and no mechanics of phenomena: "Expression in Nature is never a final symbolization, but always, and everywhere, a causal *explication*" (1968/1992, 232). It is not that there is no mechanism or determinism in Spinoza for Deleuze; rather, he is pointing out that there is a physics of force and a dynamism that allows for essence to assert itself in existence via variations of the power of action.

Deleuze seeks naturalistic norms; he turns to Spinoza's insight that the human exists as the normative type par excellence (1968/1992, 277). He writes of Spinoza developing a theory of natural right from the insights of Hobbes, one that is opposed to the classical theory of natural law. The antique tradition of natural law (Cicero) advances the following theses: (1) our being can be defined by its perfection within an order of ends (we are naturally reasonable and sociable); (2) the state of nature does not precede society, but rather we live in conformity with nature in a good civil society; and (3) in this state, what is primary and unconditional are "duties": our natural powers are only potential and require an act of reason to realize them in relation to the ends they need to serve. For Deleuze, Spinoza transforms this in a specific manner, grounding everything in natural right or power (1968/1992, 258–260).

Deleuze finds in Spinoza the demand to think in terms of capacities and powers, in which "law" is identical to "right"; natural laws are thus to be conceived as *norms of power*, rather than rules of duty. The norms in question relate to the strength and the power of action of individuals. We are normative animals because we do not wish to be only the subject of chance encounters but rather wish to seek a rational organization of our natural powers that enables the cultivation and enjoyment of them. Deleuze maintains that reason and freedom are inseparable from a formative process: "Nobody is born free, nobody is born reasonable. And nobody can undergo for us the slow learning of what agrees with our nature, the slow effort of discovering our joys" (1968/1992, 262). For Deleuze, reason is involved in all the stages of our becoming ethical and normative subjects, enabling us to move from the randomness of chance encounters to common notions and adequate ideas, and so helping us make the effort to organize our encounters, including agreements and disagreements, in a more thoughtful and rational manner (1968/1992, 280). It is not a question of moving, legitimately or illegitimately, from an "is" to an "ought," since ethics is situated in the desire of our nature to become rationally motivated normative agents of life.

Deleuze's early period closes with his masterpiece, *Difference and Repetition* (1968/1994). Recent work (e.g., Voss 2011) points to the importance of Deleuze's references to Solomon Maimon's criticism of Kant that he settled for the conditions of rational experience in general when he should have pushed for a demand for the genetic account of real experience. This hooks up with Deleuze's early work on Hume and how the mind *becomes* a human nature; the guiding question of Deleuze's Hume book is: "how is the subject generated in experience?" (1953/1991). We see Deleuze's answer in the "dynamic genesis" of Chapter 2 of *Difference and Repetition*, which treats the three syntheses of time; Deleuze's philosophical interlocutors here are Hume, Bergson, and Nietzsche. To highlight Deleuze's naturalism, we will concentrate here on organic time, the synthesis of habit producing the living present. Deleuze drives down to the most basic syntheses; he shows how beneath active syntheses (thought) are passive syntheses (perception), and how beneath passive perceptual syntheses are passive organic syntheses (metabolism) (1968/1994, 70–79; see also Ansell Pearson 1999; DeLanda 2002; Protevi 2013).

Taking account of Deleuze's encounter with Guattari is an important factor in contemporary scholarship. Toscano (2006) and Welchman (2009) posit a naturalization trajectory in which the collaborative works *Anti-Oedipus* (Deleuze and Guattari 1972/1984) and *A Thousand Plateaus* (Deleuze and Guattari 1980/1987) jettison the Kantian scaffold of subjective syntheses of time – even granting that Nietzsche is used in *Difference and Repetition* to call into question the unity of the subject.

The key term of *Anti-Oedipus* is "desiring-production," which reveals Deleuze and Guattari's conceptual and terminological innovation in their naturalism. Crisscrossing Marx and Freud, they put desire in the eco-social realm of production and production in the unconscious realm of desire. Rather than attempting to synthesize Marx and Freud in the usual way – that is, by a reductionist strategy that either (1) operates in favor of Freud, by positing that the libidinal investment of social figures and patterns requires sublimating an original investment in family figures and patterns (i.e., Oedipal triangulation), or (2) operates in favor of Marx by positing neuroses and psychoses as mere superstructural byproducts of unjust social structures – Deleuze and Guattari call desiring-production a "universal primary process" underlying the seemingly separate natural, social, and individual realms. Desiring-production is thus not anthropocentric; it is the very heart of the world. All natural processes, even those well beyond the human, are processes of desiring-production: "everything is a machine. Celestial machines, the stars or rainbows in the sky, alpine machines…nature as process of production" (1972/1984, 2). Desiring-production has two aspects relevant to our discussion of ON: (1) there is no subject that lies behind the production, that performs the production; and (2) the "desire" in desiring-production is not oriented to making up a lack, but is purely positive (1972/1984, 25). Desiring-production is autonomous, self-constituting, and creative: it is the *natura naturans* of Spinoza or the will-to-power of Nietzsche.

In *A Thousand Plateaus*, "desire" drops out of the description of nature, which is described in terms of "abstract machines" and "machinic assemblages." For Deleuze and Guattari, "machinism" is opposed to "mechanism": "machinism" denotes the creative self-organization of material systems, whereas "mechanism" denotes deadened, routinized repetition. In fact, we could say mechanism is a residue of machinism: creativity comes first, then routinization. In *A Thousand Plateaus*' terminology, "strata" (forms that induce mechanical repetition) are ontologically secondary to "lines of flight" (which provide the occasion of creative novelty by disrupting – "destratifying" and "deterritorializing" – stratified, mechanical

processes). Deleuze and Guattari write: "what is primary is an absolute deterritorialization, an absolute line of flight…it is the strata that are always residues…The question is not how something manages to leave the strata but how things get into them in the first place" (1980/1987, 56).

Despite the ontological priority of lines of flight, stratification is chronologically "simultaneous" with destratification and is a "very important, inevitable phenomenon that is beneficial in many respects and unfortunate in many others" (1980/1987, 40). Nature as process, *natura naturans*, is thus bivalent, constituting an "abstract machine" of stratification – a tendency to hierarchically ordered, mechanically repetitive systems – and destratification – a tendency to experimental, creative processes or "lines of flight." Nature as stratification is called "the judgment of God" (1980/1987, 40), while destratification (allowing creative novelty) is called "life" (1980/1987, 336, 503, 507). While stratification produces a body composed of homogenous layers, destratification allows the construction of "consistencies" or "assemblages," functional wholes that preserve the heterogeneity of their component parts and enable further nonhierarchical or "rhizomatic" connections (1980/1987, 505). The "abstract" part of the term "abstract machine" simply means that the processes of stratification and destratification occur in many material registers, from the geological through the neural, the biological through the social. An "abstract machine" is thus the diagram for processes that form functional wholes in different registers (1980/1987, 510–514). In sum, nature forms strata and also breaks down such strata, freeing parts to form connections with heterogeneous others in consistencies or assemblages.

To conclude this section, let us analyze the treatment of Deleuzean naturalism in Caygill (1997). Caygill criticizes Deleuze for sentimentally including a human level selection (active joyful encounters) and avoiding Darwin's inhuman selection. A common fear, as we see here in Caygill, is that we thus brutalize ethics and politics by naturalizing. But this is a monolithic view of nature. Darwin is not just about competition; as Kropotkin shows, there's plenty of cooperation in nature as well (Gould 1997). We can approach the intertwining of competition and cooperation in nature with Deleuze and Guattari's distinction of stratification and consistency. Nature is multiple, so both sides of human nature, the destructive and constructive, the State and the nomad, the creative and the repetitive (obviously, these do not always line up) are natural.

Antinaturalism in Critical Social Theory

Any discussion of critical social theory in the continental philosophy tradition should mention, albeit briefly, the attempts at a Marx and Freud synthesis by Reich, Marcuse, Adorno, and others. Their antinaturalism is at base a critique of the way fascists indulged in a reductive, biologizing naturalization of national characteristics as rooted in "blood" and so on. These thinkers are, however, sympathetic to a form of naturalism that consists in looking at the social and unconscious psychological factors beneath political allegiances, as in Reich (1933/1980; see also Corrington 2003), Marcuse (1955/1987), and Adorno et al. (1950/1983).

Just as we cannot fully explore the classical critical theory positions, so we must acknowledge that it exceeds the format of a survey article to discuss the antinaturalism of Western feminism; what feminist, from Wollstonecraft (1792/2009) to Beauvoir (1949/2010) and beyond, has not had to fight opponents wanting to render "natural" the causes of and justifications for women's subordinate status? We will conclude then by moving past feminism

to briefly address some key antinaturalist positions in continental philosophy-inflected critical race theory (Alcoff 2006), gender theory (Butler 1990), and disability studies (Tremain 2001). Each of these is founded on a critique of previous ideological naturalizing of what they consider to be historical, sociopolitical, contingent "constructions."

Linda Alcoff's *Visible Identities* (2006) provides a clear example of anti-MN, but perhaps a weak ON, in critical race theory, insofar as she insists on the social reality of race even as she rejects any biological basis. For Alcoff, naturalized racial identities have been exposed as specious; but the phenomenology of racial perception allows us to speak of its reality: "in the very midst of our contemporary skepticism toward race as a natural kind stands the compelling social reality that race, or racialized identities, have as much political, sociological, and economic salience as they ever had" (2006, 181). Alcoff prefers contextualism to nominalism and essentialism. In Alcoff's version of contextualism, "race is socially constructed, historically malleable, culturally contextual, and reproduced through learned perceptual practices…It can acknowledge the current devastating reality of race while holding open the possibility that present-day racial formations my change significantly or perhaps wither away" (2006, 182). Alcoff goes on to define subjectivist contextualism, which she prefers to objectivist (social science) accounts: "subjectivist approaches that begin from the lived experience of racialization can reveal how race is constitutive of bodily experience, subjectivity, judgment, and epistemic relationships" (2006, 183). Alcoff sees a danger in dehistoricizing phenomenological descriptions at the base of subjectivist contextualism: "one might be led to think that racial categories are the understandable result of the need to group and categorize. In other words, racism is the unfortunate but inevitable result of human cognitive processes" (2006, 184). Against this abstraction to the level of the "human," Alcoff insists on the historical specificity of racial perception: "although racial classification does operate on the basis of perceptual difference, it is also the case that, as Merleau-Ponty argues, perception represents sedimented contextual knowledges" (2006, 184).

In the field of gender and/or queer theory, Judith Butler is staunchly antinaturalist, if by that one means the denial of the claim that gender can be discussed in natural scientific terms (hence, Butler is anti-MN) or that gender is a natural category (she is also anti-ON). In fact, her groundbreaking *Gender Trouble* describes itself as a "critical genealogy of the naturalization of sex and of bodies in general" (Butler 1990, 147; see also Butler 1988). In *Gender Trouble*, Butler follows Foucault's *History of Sexuality, Volume 1* (1978), as she claims that "sex" as supposedly natural is in fact secreted from the social and political practices that make up the "deployment of sexuality." Butler writes, "As such, sexuality is understood by Foucault to produce 'sex' as an artificial concept which effectively extends and disguises the power relations responsible for its genesis" (1990, 92). Adding her performativity analysis to this, Butler claims that supposedly natural gender is in fact the retrospective effect of performing gender: "gender is an identity tenuously constituted in time, instituted in an exterior space through a *stylized repetition of acts*" (1990, 140; emphasis in original). Butler's performative antinaturalism is crystal clear in this formulation: the "sedimentation of gender norms produces the peculiar phenomenon of a 'natural sex' or a 'real woman' or any number of prevalent and compelling social fictions" (1990, 140).

Shelley Tremain (2001) picks up the Foucaultian term "governmentality" in her contribution to disability studies. "Governmentality" is the "action on the action of others"; it seeks to shape the field of thinkable and doable actions rather than directly manipulating or training bodies as in sovereign and disciplinary power, respectively. Tremain's target is the "social model" of disability, which naturalizes "impairment" in order to place "disability" in

social structures. Thus, in the social model, a polio-stricken leg would be an impairment made into a disability by the lack of appropriate social environment (e.g., wheelchair ramps). Building on Butler and Foucault, Tremain shows, however, that a supposedly natural impairment is itself a social category, just as "sex" is produced as precultural by the very social practices of the deployment of sexuality. Tremain writes with regard to the UK Disability Living Allowance policy: "in order to make individuals productive and governable…the policy actually contributes to the production of the 'subject of impairment' that it is claimed to merely recognize and represent" (2001, 631). The problem with naturalized "impairment" is that it presupposes a socially determined view of "human function and structure, competency, intelligence, and ability" (2001, 632). The individualizing information that social workers coerce disabled subjects into providing in order to become eligible for aid feeds them into a normalization process: "the more individualizing the nature of the state's identification of us, the farther the reach of its normalizing disciplinary apparatus in the administration of our lives" (2001, 633). The political takeaway for Tremain's antinaturalist position is then: "we should, in other words, formulate demands in terms of 'what we want,' not 'who we are'" (2001, 635).

To conclude this section, and our essay, note that one of Butler's references in *Gender Trouble* is Anne Fausto-Sterling (1989; see also Fausto-Sterling 2000; 2012), to whom she turns for biologically informed complications of the search for biological determinants of sex and gender (Butler 1988, 109). The new biology to which Fausto-Sterling contributes is a good note on which to conclude this chapter, as its notions of epigenetics and plasticity put the capacity for "social construction" into human nature (among many others, see Oyama 2000; West-Eberhard 2003; Jablonka and Lamb 2005; Wexler 2006; an introductory survey is included in Protevi 2009; 2013). Though continental philosophy-inflected critical social theory always retains some sort of social construction thesis, the new biology thinkers allow a thought of biosociality, a deep embodiment of experience. Thus, the construction is not merely a discursive or surface effect but is a construction of the material body itself in its neuroendocrinological depths. The terms "nature and culture" and "nature and nurture" need no longer be opposed; a slogan might be that "we have evolved to be so open to nurture that it becomes second nature."

Conclusion

Phenomenologists are anti-MN and anti-ON. For them, that which allows for the meaningful appearance of the world (e.g., transcendental consciousness for Husserl, Dasein for Heidegger, the lived body for Merleau-Ponty) is not a thing of the natural world, not even an emergent thing (anti-ON); it is not accessible by any natural science (even biology or psychology, let alone physics), but is the ground of any and all sciences, and therefore its philosophical articulation is independent of the natural sciences (anti-MN).

Affirmative naturalists hold to a weak MN and a weak-to-strong ON. For them, consciousness is a natural occurrence (weak ON), though it may not carry much explanatory weight (many of the affirmative naturalists are thus close to being outright epiphenomenalists, a strong ON position); what is important is an affirmative naturalist ethics that would be compatible, though perhaps not totally continuous, with natural science (weak MN).

Critical social theory is antinaturalist in refusing to countenance claims that race, gender, or disability can be discussed in natural scientific terms (anti-MN) or that race, gender,

or disability are biological categories (anti-strong ON), though it is possible to say that it has a weak ON position in recognizing race, gender, and disability as social realities.

Acknowledgments

We would like to thank Patrick Gamez, Robin James, Rebecca Millstein, Michael Rooney, Carl Sachs, and Annika Thiem for advice and encouragement.

References

Adorno, T., Frenkel-Brunswik, E., Levinson, D.J., and Sanford, R.N. (1950/1983). *The Authoritarian Personality,* abridged. New York: Norton.

Alcoff, L. (2006). *Visible Identities*. New York: Oxford University Press.

Ansell Pearson, K. (1997). *Viroid Life: Perspectives on Nietzsche and the Transhuman Condition*. London: Routledge.

Ansell Pearson, K. (1999). *Germinal Life: The Difference and Repetition of Deleuze*. London: Routledge.

Ansell Pearson, K. (2005). Bergson's Encounter with Biology. *Angelaki* 10.2: 59–72.

Beauvoir, S. de (1949/2010). *The Second Sex*, translated by C. Borde and S. Malovany-Chevallier. New York: Knopf.

Bergson, H. (1935/1977). *The Two Sources of Morality and Religion*. Notre Dame: University of Notre Dame Press.

Bergson, H. (1911/1998). *Creative Evolution*, translated by Arthur Mitchell. New York: Dover.

Butler, J. (1988). Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal* 40.4: 519–531.

Butler, J. (1990). *Gender Trouble*. New York: Routledge.

Caygill, H. (1997). The Topology of Selection: The Limits of Deleuze's Biophilosophy. In *Deleuze and Philosophy: The Difference Engineer*, edited by Keith Ansell Pearson. London: Routledge.

Corrington, R. (2003). *Wilhelm Reich: Psychoanalyst and Radical Naturalist*. New York: Farrar, Strauss and Giroux.

Cox, C. (1999). *Nietzsche: Naturalism and Interpretation*. Berkeley: University of California Press.

DeLanda, M. (2002). *Intensive Science and Virtual Philosophy*. London: Continuum.

Deleuze, G. (1953/1991). *Empiricism and Subjectivity: An Essay on Hume's Theory of Human Nature*, translated by Constantin Boundas. New York: Columbia University Press.

Deleuze, G. (1968/1992). *Expressionism in Philosophy: Spinoza*, translated by Martin Joughin. New York: Zone Books.

Deleuze, G. (1968/1994). *Difference and Repetition*, translated by Paul Patton. New York: Columbia University Press.

Deleuze, G. (1969/1990). *The Logic of Sense*, translated by Mark Lester with Charles Stivale. New York: Columbia University Press.

Deleuze, G. (1970/1988). *Spinoza: Practical Philosophy*, translated by Robert Hurley. San Francisco: City Lights.

Deleuze, G. and Guattari, F. (1972/1984). *Anti-Oedipus*, translated by Robert Hurley, Mark Seem, and Helen R. Lane. Minneapolis: University of Minnesota Press.

Deleuze, G. and Guattari, F. (1980/1987). *A Thousand Plateaus*, translated by Brian Massumi. Minneapolis: University of Minnesota Press.

Fausto-Sterling, A. (1989). Life in the XY Corral. *Women's Studies International Forum* 12.3.

Fausto-Sterling, A. (2000). *Sexing the Body*. New York: Basic Books.

Fausto-Sterling, A. (2012). *Sex/Gender: Biology in a Social World*. New York: Routledge.

Foucault, M. (1978). *The History of Sexuality, Volume 1*, translated by Robert Hurley. New York: Vintage.

Friedman, M. (2000). *A Parting of the Ways: Carnap, Cassirer, and Heidegger*. Chicago: Open Court.

Gould, S.J. (1997). Kropotkin was no crackpot. *Natural History* 106: 12–21.

Hatab, L.J. (2004). *Nietzsche's Life Sentence*. London: Routledge.

Hatfield, G. (2012). Psychology. In *The Cambridge History of Philosophy in the Nineteenth Century (1790–1870)*, edited by Allen Wood and Songsuk Susan Hahn. Cambridge: Cambridge University Press.

Hayden, P. (1998). *Multiplicity and Becoming: The Pluralist Empiricism of Gilles Deleuze*. New York: Peter Lang.

Hume, D. (1740/1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Husserl, E. (1900–01/1973). *Logical Investigations*, translated by J.N. Findlay. London: Routledge.

Husserl, E. (1910/1965). Philosophy as Rigorous Science. In *Phenomenology and the Crisis of Philosophy*, edited by Quentin Lauer. New York: Harper & Row.

Husserl, E. (1954/1970). *The Crisis of European Sciences and Transcendental Phenomenology*, translated by David Carr. Evanston: Northwestern University Press.

Jablonka, E. and Lamb, M. (2005). *Evolution in Four Dimensions*. Cambridge, MA: MIT Press.

Jonas, H. (2003). *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston: Northwestern University Press.

Kusch, M. (2014). Psychologism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/spr2014/entries/psychologism/ (last accessed July 15, 2015).

Lange, F.A. (1866/1974). *The History of Materialism*, translated by E.C. Thomas. New York: Arno Press.

Lawlor, L. (2002). *Derrida and Husserl*. Bloomington: Indiana University Press.

Lawlor, L. (2010). Becoming and Auto-Affection (Part II): Who Are We? *Graduate Faculty Philosophy Journal* 30.2: 219–237.

Lefebvre, A. and White, M. (eds.) (2012). *Bergson, Politics, and Religion*. Durham: Duke University Press.

Leiter, B. (2002). *Nietzsche on Morality*. London: Routledge.

Luft, S. and Capeillères, F. (2010). Neo-Kantianism in Germany and France. In *The History of Continental Philosophy*, vol. 3, edited by Keith Ansell-Pearson and Alan Schrift. Durham: Acumen Publishing.

Makkreel, R. and Luft, S. (eds.) (2010). *Neo-Kantianism in Contemporary Philosophy*. Bloomington: Indiana University Press.

Marcuse, H. (1955/1987). *Eros and Civilization*. New York: Routledge.

Moore, G. (2002). *Nietzsche, Biology and Metaphor*. Cambridge: Cambridge University Press.

Mohanty, J.N. (1984). Husserl, Frege and the Overcoming of Psychologism. *Philosophy and Science in Phenomenological Perspective. Phaenomenologica* 95: 143–152.

Moran, D. (2008). Husserl's transcendental philosophy and the critique of naturalism. *Continental Philosophy Review* 41: 401–425.

Nietzsche, F. (1886/1966). *Beyond Good and Evil,* translated by Walter Kaufmann. New York: Vintage.

Nietzsche, F. (1887/1967). *On the Genealogy of Morals*, translated by Walter Kaufmann. New York: Vintage.

Nietzsche, F. (1887/1974). *The Gay Science*, translated by Walter Kaufmann. New York: Vintage.

Nietzsche, F. (1906/1968). *The Will to Power*, translated by Walter Kaufmann and R.J. Hollingdale. New York: Vintage.

Overgaard, M. (2004). On the Naturalizing of Phenomenology. *Phenomenology and the Cognitive Sciences* 3: 365–379.

Oyama, S. (2000). *The Ontogeny of Information*. Durham: Duke University Press.

Petitot, J., Varela, F.J., Pachoud, B., and Roy, J.-M. (eds.) (1999). *Naturalizing Phenomenology*. Stanford: Stanford University Press.

Protevi, J. (2009). *Political Affect*. Minneapolis: University of Minnesota Press.
Protevi, J. (2013). *Life, War, Earth*. Minneapolis: University of Minnesota Press.
Reich, W. (1933/1980). *The Mass Psychology of Fascism*. New York: Farrar, Strauss and Giroux.
Richardson, J. (2004). *Nietzsche's New Darwinism*. Oxford: Oxford University Press.
Stack, G. (1991). Kant, Lange, and Nietzsche: Critique of Knowledge. In *Nietzsche and Modern German Thought*, edited by Keith Ansell Pearson. New York: Routledge.
Toscano, A. (2006). *The Theatre of Production*. Basingstoke: Palgrave Macmillan.
Tremain, S. (2001). On the Government of Disability. *Social Theory and Practice* 27.4: 617–636.
Varela, F.J. (1996). Neurophenomenology: A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies* 3.4: 330–349.
Voss, D. (2011). Maimon and Deleuze: The Viewpoint of Internal Genesis and the Concept of Differentials. *Parrhesia* 11: 62–74.
Welchman, A. (2009). Deleuze's Post-Critical Metaphysics. *Symposium* 13.2: 25–54.
West-Eberhard, M.J. (2003). *Developmental Plasticity and Evolution*. New York: Oxford University Press.
Wexler, B. (2006). *Brain and Culture*. Cambridge, MA: MIT Press.
Wolfe, C.T. (2010). Do organisms have an ontological status? *History and Philosophy of the Life Sciences* 32 (2–3): 195–232.
Wollstonecraft, M. (1792/2009). *A Vindication of the Rights of Woman*. New York: Norton.
Zahavi, D. (2004). Phenomenology and the Project of Naturalization. *Phenomenology and the Cognitive Sciences* 3: 331–347.
Zahavi, D. (2010). Naturalized Phenomenology. In *Handbook of Phenomenology and Cognitive Science*, edited by Shaun Gallagher and Daniel Schmicking, pp. 3–19. Dordrecht: Springer.

4

The Naturalism Question in Feminism

ÁSTA SVEINSDÓTTIR

Introduction

The Age of Aquarius is over. Despite the popularity of supernatural phenomena in film and television, we live in an era when a naturalist worldview is dominant in the academy. While such a worldview rejects spiritual phenomena, it is quite unclear what else it entails. In fact, on closer inspection, what goes by the label "naturalism" often turns out to be whatever the particular philosopher takes for granted that is consistent with a belief in the value of scientific enquiry. This might incline us to think there is no real question, nothing at stake, in a debate over naturalism. But we would be mistaken. Not only is there a real question; it is exceedingly important – especially for philosophical projects that ultimately have a practical aim, such as feminist philosophy.

While feminist philosophy is not a homogenous field and what goes on under the label "feminist" is as diverse as what gets labeled "naturalist," it has a single practical aim, which is to end the oppression of women. Feminists and philosophers committed to ending other kinds of oppression are natural allies, and many feminists believe that working against sexist oppression can only be done in conjunction with working against other kinds of oppression. Others focus exclusively on sexist oppression. That does not matter for our discussion here. The naturalism question is equally important for all antioppression projects.

What is the naturalism question in feminism? It is this:

To what extent should a feminist embrace naturalist commitments?

This essay is dedicated to clarifying what is at stake in answering this question. My aim will be to isolate naturalist claims that are contested within feminism; to isolate, if you will, the naturalism question in feminism. This will not be a comprehensive survey of various

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

feminist philosophers' stances toward naturalism, but rather a critical engagement with contested naturalist commitments from a feminist perspective.[1]

We begin by discussing what is meant by "naturalist commitments."

Naturalist Commitments

While most feminists who embrace naturalism cite Quine as their inspiration,[2] we do well to start with Hume on induction[3] to get a sense of what the key naturalist commitments come to. After all, the position Quine advocated had its roots in Hume, or, as Quine put it, "the Humean predicament is the human predicament."[4] What I shall do is tell a broadly Humean story about the naturalization of normativity. My very brief story begins with Hume's skepticism about inductive inference and generalizes to other phenomena where one might expect to find a correct/incorrect distinction, including other epistemic phenomena, but also moral, social, aesthetic, and so on. I don't offer this as the historically correct view of Hume; it may be a portrait of an ancestor who exists only in our family lore.

Hume argued[5] that we cannot justify rationally our inductive inferences. For instance, there is no rational principle that can get us from "in human memory, the sun has risen every morning" to "tomorrow the sun will rise," and hence when we draw that inference, we do not do so on the basis of a *rational* ground, but on a *nonrational* one, namely, habit or custom.

1 Many feminist philosophers have advocated naturalism in their subfield of feminist philosophy. Prominent here are Louise Antony (1994; 2000), Lorraine Code (1987; 1991; 1995; 1996), Lynn Hankinson Nelson (1990; Hankinson Nelson and Nelson 2003), and Elizabeth Potter (1995; 2001) in epistemology, Helen Longino (1990) in philosophy of science, Alison Wylie (1999) in philosophy of archaeology, and Alison Jaggar (2000) and Margaret Urban Walker (2001) in ethics.

2 Cf. Antony (1994), Jaggar (2000), and Walker (2001).

3 Hume (1748, §§ iv and v).

4 Quine (1969, 69).

5 The textbook interpretation goes like this: We are looking for the justification of our belief that the sun will rise tomorrow, given that it has risen every day in human memory. We rely on the sun's rising; what is our reliance based on? How do we get from:

(1) In human memory, the sun has risen every morning

to

(C) The sun will rise tomorrow

On Hume's view, statements come in two kinds: Relations of Ideas and Matters of Fact. Our belief in the truth of the latter is justified by an appeal to an empirical investigation of some sort, be it perception or testimony, or the conclusion of an argument with premises justified in that way. Our belief in a Relation of Ideas is justified by ascertaining that the negation of it yields a contradiction.

(1) is a Matter of Fact, and our belief in it is justified by an appeal to our own experience and the testimony of others. But how do we get from (1) to (C)? We need the conditional, (2), to get there, but how is (2) justified?

(1) In human memory, the sun has risen every morning

(2) If in human memory, the sun has risen every morning, then the sun will rise tomorrow

(C) The sun will rise tomorrow

(2) is not a Relation of Ideas, as its negation does not yield a contradiction, and it is not an ascertainable Matter of Fact, so our belief in it must rely on our belief in another statement. (3) will do:

(3) The future will resemble the past

But now we ask the same question: How is our belief in (3) to be justified? It is neither an ascertainable Matter of Fact, nor a Relation of Ideas, so our belief in it must rely on our belief in yet another statement. (4) will do:

(4) In the past, the future has resembled the past, so in the future, the future will resemble the past

And now we are in a pickle. Our belief in (4) is justified by our belief in (3), so (4) cannot underwrite our belief in (3). Our justification is circular.

It would be mistaken to say that induction is *justified* by reference to habit or custom. That we are in the habit of doing something cannot ground our claim to its *rightness*. Custom may offer a *causal explanation* of how we came to have a belief, but it is not a *rational justification* for that belief. Does this then mean that we cannot say that some inductive inferences are right and others wrong? Yes, in the absence of a further story that justifies our habits, we cannot say that one inference is better than another. We can only say that we draw certain inferences, not others. From now on, I will reserve the word "justification" for the *normative* notion of that which grounds our claim to the rightness of our beliefs or actions. We can then say that the upshot of Hume's argument is a naturalism about the normativity of inductive inferences: there is no such thing as the right or wrong inferences; there is no epistemic normativity in the domain of inductive inferences.

Now that it is clear what naturalism about normativity with respect to inductive inferences is, we can generalize from this reading of Hume to what I call "naturalism about normativity" about other epistemic phenomena, be they practices, actions, or attitudes. We can then further generalize from the epistemic to other kinds of actions, practices, and attitudes, such as moral, political, social, or aesthetic:[6]

Naturalism about normativity with regard to X (where X is an action, attitude, or practice):
There is no such thing as normativity with regard to X and no such thing as a justification for X: there aren't right and wrong practices, better and worse ones, just ones we engage in and ones we don't. Likewise, there is no justification for a particular action or attitude, just a causal explanation of it.

A philosopher may be a naturalist about all normativity, or only certain kinds of normativity. On my reading of Hume, Hume is a naturalist about epistemic norms relating to inductive inferences, in the sense that traditionally epistemic practices, including inductive inferences, had been thought to be grounded in reason, and certain inductive inferences thought to be better justified rationally than others. Hume thinks that there is no such epistemic normativity, as the traditional understanding would have it. And if there isn't, how are we to think of the task of epistemology? What are epistemologists left to do? We can do an empirical investigation into how we form beliefs – what our epistemic habits are – and epistemology then becomes a branch of psychology. This is precisely what the 20th-century naturalist W.V.O. Quine advocated in the influential essay "Epistemology Naturalized" (1969).

Naturalism in Quine's hands is characterized by a rejection of normativity and commitment to epistemology as an empirical science:

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject…But a conspicuous difference between old epistemology and the epistemological enterprise in this new psychological setting is that we can now make free use of empirical psychology. (1969, 82–83)

Our main focus in this chapter will be on the stance toward normativity, but near the end we will discuss briefly the role of empirical enquiry in philosophical theorizing.

6 That naturalism and the scientific worldview have no room for normativity is a common theme. Cf. De Caro and Macarthur (2010); for early formulations, see Kim (1988) and Kornblith (1994); for feminist discussions see, for example, Clough (2004), Janack (2004), and Tanesini (2010).

Feminist philosophers often focus on the question of normativity with respect to a particular domain or practice. But even those who advocate naturalism about normativity in ethics or political philosophy typically do so by appealing to the epistemic case (Jaggar 2000; Walker 2001). Naturalized epistemology thus provides the starting point for feminists who advocate naturalist approaches in other subfields of philosophy.

There are other theses that a commitment to naturalism might involve, but what I have called "naturalism about normativity" is the sticking point for feminists. It is both that naturalism about normativity seems to promise powerful critical resources in the fight against sexist oppression, because the justification for oppressive institutions and practices is undercut, and that it leaves feminists with few resources with which to construct a theory of what makes an action or practice right or wrong.

Let us see what naturalism about normativity in its crudest form can offer to a critical project of fighting sexist oppression, and why it falls short of supporting a normative theory.

Naturalism about All Normativity in Feminist Hands

Feminist philosophers have been attracted to the critical potential of naturalism about normativity, and for good reason. In broadest terms, a critical move that relies on naturalist commitments is part of an ideology critique: the justificatory story that is offered to underwrite sexist attitudes, actions, practices, or institutions is undercut. The structure of the criticism is this: A claim to the effect that a certain attitude, action, or practice is justified is undercut by showing it to lack such justification, because hope for justification in that realm of human inquiry and conduct is misplaced. The claim of *rightness* is undercut, and thus the justificatory story – the ideology – is rendered null and void[7] and the process of liberation from such oppressive attitudes, actions, or practices can begin.

But even though naturalism about normativity with regard to a certain phenomenon is a powerful critical tool, and is especially useful for feminists engaged in ideology critique, its critical potential vanishes for the philosopher who rejects all normativity. This is an important point and worth going through in detail.

Consider a form of ideology critique that involves undercutting the story used by an employer to justify a certain hiring for positions. This is a critique that is epistemic, as opposed to moral, in that the claim at issue is a claim about the employer's judgment of the competence of candidates. In this scenario, the employer judges a male candidate to be better qualified than a female one, based on the evaluation of their CVs. The feminist critiques the employer's claim that the male is more competent than the female by referring to studies of implicit bias. These studies have shown that when one candidate is found more competent than another, it is often (subconsciously) based on various associations employers have with the markers on the candidates' resumes. For example, studies have shown that people evaluate CVs with a male name at the top more favorably than an identical CV with a female name and are more likely to hire the male (Steinpreis, Anders, and Ritzke 1999).[8] While the employer is *caused* to

7 I am using "ideology" here to designate the justificatory story offered to underwrite a phenomenon, be it a practice, attitude, or what have you. I do not mean to use it as a pejorative. While many justificatory stories are pernicious and mask an unjust arrangement, some are not and do not.

8 This kind of bias in hiring is documented, and there is now a large body of literature on implicit bias. Most of the studies have focused on implicit bias related to gender or race. The results of the first study were published in Greenwald and Banaji (1995).

favor one candidate over another, the employer is not rationally justified in doing so; so the claim that the employer hired the best-qualified candidate is undercut. They hired the one they were psychologically compelled to hire. The claim that the hiring is *right* is undercut, and so is the ideology accompanying the hiring.

But what exactly have we accomplished in this manner? In what sense is the ideology "undercut"? Did we show that the ideology was *incorrect*? And more to the point, did we show that the hiring was therefore *wrong*?

For the hiring to be wrong, it minimally has to be that the judgment of competence was wrong. The latter is an epistemic matter, the former moral. Do we have any ground to stand on if we want to claim that the hiring was *wrong*? The structure of the argument is this:

(1) If there is no epistemic justification in the realm of hiring, then it's not the case that a judgment of competence is epistemically right.
(2) If it's not the case that a judgment of competence is epistemically right, then it's not the case that the hiring was morally right.
(3) There is no epistemic justification in the realm of hiring.
(4) It is not the case that the hiring was morally right.

But the following argument works just as well:

(1) If there is no epistemic justification in the realm of hiring, then it's not the case that a judgment of competence is epistemically wrong.
(2) If it's not the case that a judgment of competence is epistemically wrong, then it's not the case that the hiring was morally wrong.
(3) There is no epistemic justification in the realm of hiring.
(4) It is not the case that the hiring was morally wrong.

The real critical bite of naturalism about normativity here thus seems to be not that a certain hiring practice was biased or wrong, but that the claim to its being the right hiring was wrong. According to the feminist critic, the employer is not *entitled to claim* that the hiring is *right*. They are not entitled to it, because there is no such thing. Hirings are not the kind of thing that can be right or wrong. It is a category mistake to say that a hiring is right.

But isn't the aim of the feminist to criticize the hiring practices, and not just the discourse about them? For that is all we seem able to do: criticize discourse. We then become the talk police: "You shouldn't say this or that kind of thing!" And why not? You would be committing the gravest sin of all: the category mistake!

Matters are worse still for the feminist who rejects all forms of normativity. For how would we even be able to engage in the ideology critique at all? How could we even accuse the employer of committing a category mistake? How could we assert that the employer is wrong in claiming that the hiring was justified? The critique of the employer, which involves claiming that it is wrong because it is a category mistake, goes like this:

(1) Employer claims that a is F.
(2) It's not the case that a is F, because asserting that a is F is a category mistake.
(3) Employer is in error in claiming that a is F.

The kind of normativity the feminist critic needs to make this argument is minimal, but is still a species of normativity. We need the notion of a category mistake, the principle that one is in error when one asserts the false, and simple propositional logic.

(1) Employer asserts that a is F.
(2) If S asserts P, and −P, then S is in error in asserting P [one is in error when one asserts the false].
(3) −(a is F) and −(a is −F) [category mistake].
(4) Employer is in error in asserting that a is F [from (1), (2), (3)].

The main point of interest here is the principle that one is in error when one asserts the false. This is not to be confused with norms of assertion, Gricean or otherwise (cf. Grice 1989), for it is not about whether the speaker is praiseworthy or blameworthy, but rather whether they are correct or incorrect in their assertion. Do they assert the facts as they are? We obviously have to rely on the notion of truth and falsity, which not everyone will like, but we also have to take on board some standard of correctness: a contemporary of Pericles is incorrect in asserting that the earth is flat, though they are not blameworthy; in fact, given their evidence, they are entitled to assert that the earth is flat. Nevertheless they are incorrect, because the earth isn't flat.

But a philosopher who rejects all normativity ought not to like the principle that one is in error when one asserts the false. On their account, it isn't the case that there are certain assertions that are correct and others that are incorrect; there are just the ones we do make and the ones we don't. Such a philosopher not only cannot criticize the employer's hiring of the candidate, but cannot engage in discourse criticism either, since rightness and wrongness with regard to claims and assertions are misplaced. But the philosopher who rejects all forms of normativity cannot even make sense of the notion of misplacement, since that notion presupposes that there are things that are *correctly* placed, and others that are misplaced. By hypothesis, the philosopher who rejects all forms of normativity cannot make sense of a correct/incorrect distinction, and thus this line of argument isn't available to them. The naturalist worldview, in which there are no norms, thus depicts a world in which there is no legitimate criticism either. It is also a world without injustice and oppression, since it is a world entirely devoid of value. Meanwhile, pain and suffering continue.

Obviously, it isn't by itself a good argument against naturalism about all normativity that, if we feminists embrace it, then we don't have any critical tools by which to argue for a better world. But this is as good a point as any at which to linger with the aim of feminist theorizing and with what the methodology thus has to be. We start with the widespread lived experience of oppression and mistreatment and look for theories that can explain such experience and the suffering that accompanies it. Such theories are inherently normative, since the naturalistic worldview leaves no room for harm and suffering. Only in the face of very strong arguments to the effect that our phenomenology is in radical error (which include the claim that we are not suffering after all, contrary to our immediate evidence) should we abandon our attempt to make sense of that phenomenology and seek theories that can make the situation better. But no argument has been presented that would force us to abandon our quest. And that is why we feminists should reject a worldview that banishes all normativity.

Evolutionary psychology (cf. Driscoll 2013) provides interesting case studies, given the foregoing discussion. According to evolutionary psychology, social phenomena are caused

by biological phenomena. For instance, the prevalence of domestic violence is causally explained by reference to genetic factors in males, and the gendered division of labor is explained causally by reference to biological differences between males and females. Apart from the various other theoretical criticisms that can be levied against such views – not least of which is the availability of superior alternative explanations (cf. Fehr 2011) – we note a striking feature of such explanations. They are offered as justificatory stories (ideologies accompanying practices) in the face of criticism of these practices as unjust. But if the social is causally determined by the biological, there are no such things as justifications, and hence a justificatory story is unavailable. And the worldview we are offered is the aforementioned world without norms. It is a world without suffering and oppression,[9] to be sure, because nothing in it is bad or good. It is also a world that leaves us without hope that anything will ever get better. As we have seen, we have reasons to reject this worldview.

Pragmatic Normativity?

Let's go back to Hume on induction and ask: Is there any way to distinguish between the habits we engage in so we can say that some are better than others in a way that suffices for feminist aims? There are two kinds of attempt to make sense of some habits as being better than others: the first, that some are somehow more *reliable*, and the second, that they are somehow more *useful*. I'm taking the first notion to be a non-normative, naturalistic notion and thus not of use to the feminist, given the foregoing discussion. But what about the latter?

Nelson Goodman, in "The New Riddle of Induction" (Goodman 1955), follows Hume in rejecting a rational justification for inductive inferences. He then goes one step further and argues that some inferences are better than others because of their track record.[10] Certain kinds of inference have been useful in the past, while others have not. In that sense, some inferences are better than others. Importantly, we cannot say that they have a better track record because they are truer or more accurate; that would be to build an epistemic notion of correctness into the notion of *usefulness*, which we don't want to do. Goodman rejects epistemic normativity in the traditional sense, but he allows for some sort of normativity, namely a pragmatic kind.

We can follow Goodman and ask whether we can make use of pragmatic normativity to distinguish among inferences we make. Would it be sufficient for feminist purposes to be able to say of certain inferences or actions that they are more or less useful or more or less prudential for us? We would deny that there is a sense of "epistemic" that is distinct from the pragmatic and the prudential, such that it is epistemically better to draw one inference rather than another, or better to act in a certain way. But we could draw distinctions among phenomena on pragmatic grounds. Could we, for example, critique the employer's hiring if we relied only on pragmatic norms? We would then say that the employer's judgment of the candidates was not epistemically in error, but pragmatically inferior: it would have been more prudential to have judged the candidates equally qualified, given their CVs, than it was to judge them unequal. Again, we cannot explain utility in terms of a judgment's correctness; we can only rely on the track record in our claim for usefulness. This means

9 I take both *oppression* and *suffering* to be normative notions.

10 In Goodman's view, certain predicates (e.g. "green" and "blue") are more "projectible" than others ("grue" and "bleen"), because of their track record.

that if there is no track record, we have no argument. This also means that if bias is somehow prudential or useful, if it's prudent not to recognize the qualifications of a candidate, for example, that's OK (and there can be little doubt that throughout human history, bias has been useful to many). This pragmatic notion has little bite in the face of injustice: whatever is more useful is better; there is no other sense of "better" or "worse." The right is held hostage to what is useful.

From the aforementioned considerations, we can see that feminists are not helped by an appeal to a pragmatic notion of normativity, even though we might on occasion be able to give a pragmatic argument against an unjust arrangement.

Naturalism about Some Kinds of Normativity

I have given an argument that relies on the lived experience of women and others who suffer oppression against naturalism about all normativity. Philosophers who aim to end all forms of oppression have strong reasons to reject the view of the world that has no norms. This is a world that has no reasons, only causes; no justifications of any kind, only explanations. We have also seen that the feminist cannot make do with only pragmatic norms. It is now time to turn to a more tempered embrace of naturalism about normativity and its possible usefulness for feminists in particular.

A tempered naturalism might reject some forms of normativity, say inductive or aesthetic, while endorsing others, say epistemic or moral normativity, so the tempered feminist naturalist might reject some entities or alleged norms while endorsing others that are central to her case against gender injustice. Alternatively, the feminist thinker might reject naturalism *per se*, but appeal to naturalist explanations to undercut certain claims; naturalistic explanations would thus form part of an ideology critique. Let me address each in turn.

Naturalism about a Specific Kind of Normativity

Given the feminist aim of critiquing gender injustice, is it open to feminists to be naturalists about a certain kind of normativity, say epistemic or moral? Not, I claim, if we want to be able to criticize moral or epistemic wrongdoing. If we reject epistemic normativity, we cannot account for such things as epistemic injustice and incorrect judgment; if we reject moral normativity, we cannot legitimately say that it is morally bad for a husband to beat up his wife or molest his children. It is worth lingering with this point. In each case, it isn't that we are embracing a worldview devoid of value, but that we are embracing a worldview devoid of a particular kind of value, be it epistemic or moral.

Consider, for example, the employer's judgments of competence. If we want to be able to say that the employer's judgment was *in error*, that it would have been *correct* to judge the candidates equally competent, given their identical CVs, then we need epistemic normativity. And if we are persuaded by Miranda Fricker (2007) that there is a distinct type of epistemic injustice, as in the case where markers on the candidate's body or CV are associated with a stereotype of a social group that biases employers' judgments of competence, then we need to be able to legitimately say that the judgment was wrong, epistemically as well as morally.

Similarly, if we want to advocate for laws that prohibit and prosecute child molestation or sex trafficking because such behavior is morally wrong, we need moral normativity. Again, in the absence of a strong argument to the effect that the lived experiences of millions of people are somehow in error, those lived experiences are the starting point of our theorizing.

Perhaps there are domains of human life the feminist would be happy to allow are devoid of norms. Perhaps aesthetics is such a domain. We need not settle that here. But domains in which injustice against women and other oppressed groups is to be found are domains in which the feminist needs norms and so must reject naturalism about those norms.

In the final section, I will argue that naturalistic explanations can provide a heuristic device that can be used as part of an ideology critique.

A Naturalist Move as Part of an Ideology Critique

Consider some classic cases of sexist claims that feminist have fought against:

(1) Women are better at caring for children and therefore should stay at home and take care of children.
(2) Women are inferior intellectually and resources should not be spent to educate them.
(3) Women are unable to make their own choices and therefore should not be allowed to vote.
(4) Women who have children are scatterbrained and should not be hired for jobs.
(5) Battered women lack the courage to leave their husbands and deserve what they get.
(6) Women are helpless and need a man to take care of them.
(7) Girls are worse at mathematics than boys and resources shouldn't be spent to educate them in mathematics.

These claims can all be put into the form of a deductive argument, with the recommendation as the conclusion. Arguments where the premises are false are, of course, easy to refute. The interesting cases are the ones where the premises may indeed be true, however contingently so. Consider (1), for example:

(a) Women are better at caring for children.
(b) If women are better at caring for children, then they should stay at home and take care of children.
(c) Women should stay at home and take care of children.

The premise, (a), may be true in a society that confines women to the role of caring for children and in which the upbringing of little girls is all geared toward that future role. But an exploration of why women are in general better at caring for children in that society reveals the possibility that they might not have been, if only sociocultural and material circumstances were different. In particular, had they been able to develop their other talents and engage in activities outside the home in that society, they might not have developed their skills at childrearing. So, instead of justifying keeping women at home with the children, women's childrearing skills are shown to be possibly a result of that setup.

Here we have a classic case of an ideology critique: we reject a justificatory story in favor of a possible causal explanation for the phenomenon and thus undercut the ideological

basis for the arrangement in question. It is also a classic case of a naturalist move: we look at the actual situation the people are in and offer a possible causal explanation for their predicament. The causal explanation undercuts the normative claim. Unlike in the case where we reject normativity altogether, it isn't that the causal explanation undercuts the normative claim because normative claims are somehow misplaced, but rather that offering the causal explanation of the predicament in question shows that, minimally, the facts aren't all in. In some cases, we need to reserve judgment until we have controlled for all the variables; in others, empirical studies support a complete rejection of the recommendation.[11]

Here are similar ways in which the other claims can be countered:

(1) Women are found to be inferior intellectually in certain societies and contexts because resources have not been spent to educate them.[12]
(2) Women are unable to make their own choices in certain societies and contexts because they have not been allowed independence from fathers and husbands.
(3) Women who have children are scatterbrained in certain societies and contexts because they have no help with childcare and are overworked and tired.
(4) Battered women lack the courage to leave their husbands in certain societies and contexts because they lack social and economic support to do so.
(5) Women are helpless in certain societies and contexts because they are prevented from developing their talents.
(6) Girls are worse at mathematics than boys in certain societies and contexts because they are prevented materially and socially from developing their mathematical talents.

In each case, there is a corresponding demand to rectify the situation. For example, there is a demand that the society in question provide support and education for women to develop their talents and foster their independence; also a demand that society make available childcare and other kinds of support for mothers so they can participate fully in that society.

These responses always turn on offering a causal explanation (and in some cases, an alternative explanation) of the truth of the premises of the argument that was offered, one that undercuts the justification of some arrangement, practice, right, or privilege (or lack thereof). It is a naturalist move because we counter a normative claim with a causal explanation, but it is done as part of an ideology critique and not because the normative concern is somehow misplaced. For instance, in the case of the claim that girls are worse at math than boys in some societies, not only do we offer a possible causal explanation for that truth which suggests the possibility that it might not be true if only circumstances were different, but we actually provide empirical evidence that girls not only do not lag in math in some countries but actually do better than boys. Moreover, we can offer empirical evidence that the greater gender equality in a society, as measured along several axes, the smaller the gender gap in mathematics performance (Kane and Mertz 2012).

In offering causal explanations and empirical data as part of a feminist critique, feminists clearly side with Hume and Quine in insisting that empirical methods have a role to play in philosophy. Philosophy isn't an empirical science, *pace* Quine, but empirical methods and

11 In the first kind of case, lack of resources and access and so on *might* be the explanation for the predicament; in the second kind of case, it demonstrably is. The case of difference in math performance is an example of the latter. See Kane and Mertz (2012).

12 Mary Wollstonecraft argues something like this in Wollstonecraft (1792).

questions are relevant to normative enquiry. Feminists believe that the social and material conditions and lived experiences of oppressed people are relevant to our theorizing. It is an empirical question what these are. Hence, empirical questions are relevant to our theorizing. But philosophy isn't exhausted by empirical enquiry, because empirical methods can never answer the question what it is that makes a certain attitude, action, or practice right or wrong. We need theorizing that is inherently normative to answer such a question.

Conclusion

The naturalism question in feminism is this: To what extent should a feminist embrace naturalist commitments? I have characterized naturalism as involving two commitments: a rejection of normativity and a commitment to philosophy as a descriptive discipline consisting of empirical questions to be answered by empirical methods. I have argued that a feminist should not be a naturalist about normativity, because feminists need to engage in an inherently normative enquiry. On the other hand, a naturalist move, wherein one offers a causal explanation to undercut a normative claim, is an essential part of a good ideology critique, which is a feminist staple. As feminists believe that our lived experiences and social and material conditions are relevant to our theorizing, empirical questions should play an integral role in normative enquiry, but since philosophy is not exhausted by such empirical enquiry, philosophy isn't an empirical science.

References

Antony, L. (1994). Quine as Feminist: The Radical Import of Naturalized Epistemology. In *A Mind of One's Own*, edited by L. Antony and C. Witt. Boulder: Westview Press.

Antony, L. (2000). Naturalized Epistemology, Morality, and the Real World. *Canadian Journal of Philosophy* 30(Suppl.): 103–137.

Code, L. (1987). *Epistemic Responsibility*. Hanover: University of New England Press.

Code, L. (1991). *What Can She Know? Feminist Theory and the Construction of Knowledge*. Ithaca: Cornell University Press.

Code, L. (1995). *Rhetorical Spaces: Essays on Gendered Locations*. New York: Routledge.

Code, L. (1996). What is Natural about Naturalized Epistemology? *American Philosophical Quarterly* 33(1): 1–22.

Clough, S. (2004). Having It All: Naturalized Normativity in Feminist Science Studies. *Hypatia: A Journal of Feminist Philosophy* 19(1): 102–118.

De Caro, M. and Macarthur, D. (eds.) (2010). *Naturalism and Normativity*. New York: Columbia University Press.

Driscoll, C. (2013). Sociobiology. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/sociobiology/ (last accessed July 15, 2015).

Fehr, C. (2011). Feminist Philosophy of Biology. In *The Stanford Encyclopedia of Philosophy,* edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/feminist-philosophy-biology/ (last accessed July 15, 2015).

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Greenwald, A.G. and Banaji, M.R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review* 102: 4–27.

Grice, P. (1989). *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.
Hankinson Nelson, L. (1990). *Who Knows: From Quine to a Feminist Empiricism*. Philadelphia: Temple University Press.
Hankinson Nelson, L. and Nelson, J. (eds.) (2003). *Feminist Interpretations of W.V. Quine*. University Park: Penn State University Press.
Hume, D. (1748). *An Enquiry Concerning Human Understanding*. London: A. Millar.
Jaggar, A. (2000). Ethics Naturalized: Feminism's Contribution to Moral Epistemology. *Metaphilosohy* 31(5): 452–468.
Janack, M. (2004). Feminist Epistemology. In *The Internet Encyclopedia of Philosophy*. Available from: http://www.iep.utm.edu/fem-epis/ (last accessed July 15, 2015).
Kane, J.M. and Mertz, J.E. (2012). Debunking the Myth: Sex Differences in Math Performance. *Notices of the American Mathematics Society* 59: 10–21.
Kim, J. (1988). What Is "Naturalized Epistemology"? In *Philosophical Perspectives Volume 2: Epistemology*, edited by James Tomberlin. Atascadero: Ridgeview.
Kornblith, H. (1994). *Naturalizing Epistemology*, 2nd edn. Cambridge, MA: MIT Press.
Longino, H. (1990). *Science as Social Knowledge*. Princeton: Princeton University Press.
Potter, E. (1995). Good Science and Good Philosophy of Science. *Synthese* 104: 423–439.
Potter, E. (2001). *Gender and Boyle's Law of Gases*. Bloomington: Indiana University Press.
Quine, W.V.O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Steinpreis, R., Anders, K., and Ritzke, D. (1999). The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study. *Sex Roles* 41(7/8): 509–528.
Tanesini, A. (2010). Naturalism and Normativity: Rorty and Barad. *Feminist Interpretations of Richard Rorty*, edited by Marianne Janack. University City: Pennsylvania State University Press.
Walker, M.U. (2001). Seeing Power in Morality: A Proposal for Feminist Naturalism in Ethics. *Feminists Doing Ethics*, edited by Peggy DesAutels and Joanne Waugh. Lanham: Rowman & Littlefield Publishers.
Wollstonecraft, M. (1792). *A Vindication of the Rights of Woman*. London: J. Johnson.
Wylie, A. (1999). The Engendering of Archaeology: Refiguring Feminist Science Studies. In *Science Studies Reader*, edited by Mario Biagioli, pp. 553–568. New York: Routledge.

5

On Naturalistic Metaphysics

THOMAS M. CRISP

Naturalism – on one popular understanding of the term – is the view that there is no good reason to believe in God or anything relevantly like God, and no need to appeal to such an entity in explaining or understanding the world. No part of the world, says the naturalist, whether it be the mental, the physical, or the moral parts, requires postulating non-natural entities, forces, or processes to explain and understand its workings.

Naturalism is highly fashionable in Western academia. Its journals brim with attempts to give naturalistic accounts of the mental, the moral, the physical, and all else besides. In some quarters, especially the sciences, to suggest that there are or might be non-naturalistic explanations of this or that phenomenon is to court professional disaster, and to make oneself the target of vituperative accusations of intellectual sloppiness and irresponsibility. Naturalism, in many sectors of the Western academy, has become orthodoxy.

But there is a deep and little discussed problem with this research program, a problem that threatens the coherence of the whole project. After some further explanation of what naturalism is, I will turn to an explanation of this problem.

What Is Naturalism?

Naturalism, I say, is the view that there is no good reason to believe in God (for present purposes, the omniscient, omnipotent, wholly good creator of all), or anything relevantly like God, and no need to appeal to such an entity in explaining or understanding the world. But what would it be for a being to be "relevantly like" God? A being is "relevantly like God," let us say, if it is a mind (something with beliefs, desires, aims, and intentions) whose activities explain the origin, structure, and ongoing existence of the cosmos.

I said too that, according to naturalism, no part of the world, whether it be the mental, the physical, or the moral parts, requires postulating non-natural entities, forces, or processes to explain and understand its workings. And what are "non-natural" entities,

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

forces, and processes? Entities, forces, and processes are non-natural, let us say, if they cannot be completely described without appeal to the existence of a mind whose activities explain the origin, structure, and ongoing existence of the cosmos.[1]

On Naturalistic Metaphysics

I have suggested that there is a fundamental problem with the project of explaining the mental, the moral, and the physical in naturalistic terms. It'll help in explaining that problem to say more about what that project comes to.

Appearances can be deceiving.[2] Someone might appear to be a friend but really be a foe; something might appear to be near but really be far; the earth beneath our feet appears to be stationary but is really rotating; and so forth. Sometimes it's possible to get behind misleading appearances and see things as they really are. So: when we discovered that the earth is really rotating, we got behind the appearance that the earth is stationary and saw into how things really are.

Metaphysics is the attempt to do this: it's the attempt to get behind the appearances to the reality underlying them. More carefully: it's the attempt to get behind the appearances and understand the nature and structure of the reality underlying them.

We can think of metaphysics as coming in more or less abstract varieties. To the extent that the constituents of the structures postulated by a metaphysical theory differ from the objects of everyday experience; to the extent that the relationships of dependence between the postulated structures and the world of appearances are complicated (requiring complex mathematics, say, and/or long chains of complicated reasoning to describe); to the extent that knowledge of these structures and the dependence relationships connecting them to the world of appearance is far removed from the concerns of everyday life; and to the extent that questions about these structures and dependence relationships have been the subject of perennial philosophical dispute: thereby, the theory in question is more abstract.

A relatively nonabstract form of metaphysics is the sort of inference to the best explanation hunter-gatherers use when tracking wounded animals, postulating hypotheses about the animals' injuries and locations from appearances on the trail. Such attempts to get behind the appearances and understand the nature and structure of the reality underlying them are, as we're using terms, a form of metaphysics. Since knowledge of these underlying realities and their relationships to the appearances is highly connected to concerns of everyday life, and the postulated realities and their connections to the appearances are of a familiar, everyday sort, this represents a relatively nonabstract sort of metaphysics.

Not so with, for example, string theory: here, one postulates strange particles, multidimensional spacetimes, and so forth, knowledge of which is far removed from the concerns of everyday life, and whose relations of dependence to the world of everyday

1 It's worth noting that naturalism, as I have defined it, is not the same thing as atheism. Naturalism, on my definition, says there is no good reason to believe in God; but it could be, of course, that though we have no good reason for thinking so, there nevertheless is such a being as God. So naturalism doesn't straightforwardly entail atheism. Nor does atheism straightforwardly entail naturalism. You could be an atheist, in that you deny the existence of an all-powerful, all-knowing, wholly good creator, but join Hegel, for instance, in thinking that total explanation of the world requires appeal to a cosmic mind of a nontheistic sort. The key point, then: naturalism, on my definition, is not the same view as atheism.

2 The next two paragraphs owe an obvious debt to Peter van Inwagen's (2009, vi–xi) characterization of metaphysics.

appearances are describable only with the help of extraordinarily subtle and complex mathematics. Here we have a highly abstract form of metaphysics.

Such, then, is metaphysics, and some indication of what it is for metaphysics to be more or less abstract. Naturalistic metaphysics, let us say, is metaphysics such that the structures that it postulates to explain the appearances are populated entirely by "natural" entities, forces, and processes: entities, forces, and processes completely describable without appeal to a mind or minds whose activities explain the origin, structure, and ongoing existence of the cosmos.

And such is the project of explaining the mental, the moral, and the physical in naturalistic terms: naturalistic metaphysics. It's the attempt to get behind the mental, moral, and physical appearances, to understand the nature and structure of the realities underlying them, all the while postulating only natural entities, forces, and processes. And as it occurs in the pages of philosophical and scientific journals and books, it's mostly highly abstract metaphysics. Chapter 2 in this volume is an elegant example. It sketches what it would look like to give a purely naturalistic account of moral phenomena, pointing out along the way that there are no irreducible, non-natural moral properties, that there is no such thing as metaphysical freedom or agent causation, that quantum-mechanical processes in the microtubules of certain neuronal segments would do nothing to secure the existence of agent causation, that there is no such thing as a faculty of the will, and that all moral truths are grounded in facts about human needs, desires, and purposes: these are, one and all, claims about realities far removed from the concerns of everyday life, all subject to massive and perennial philosophical dispute. Some highly abstract, naturalistic metaphysics on display here.

An Objection to Naturalistic Metaphysics

I turn now to an objection to abstract naturalistic metaphysics (for short, *naturalistic metaphysics*). It rests on three theses. First, there is this:

(1) *Evolutionary Thesis:* If naturalism is true, if there is no need to appeal to God or anything like God in explaining or understanding the world, then it is highly likely that we humans and our cognitive faculties are the products of unguided evolutionary processes of the sorts described by contemporary evolutionary theory.

Where by "unguided," I mean that the processes that gave rise to us were not orchestrated or superintended by an intelligent agent – that they were "blind," to use Richard Dawkins' (1986) term. If neither God nor any being like God figures into the explanation of how the world works, then neither figures into the explanation of how we humans and our cognitive faculties arose, and if that's so, it's extremely difficult to see how we could have got here if not by unguided processes of the sort described by contemporary evolutionary theory.

Next, there is this:

(2) *Thesis of Unreliability:* The probability that we humans command much by way of reliable insight into abstract metaphysical matters, given naturalism and that we and our faculties are the product of evolutionary processes of the sort described by contemporary evolutionary theory, is *inscrutable* (i.e., we have no way of knowing its value).

Why think this thesis true? Well, is there some reason, given naturalism and that we and our cognitive faculties arose via unguided evolutionary processes of the sort described by contemporary evolutionary theory, to expect those processes to have endowed us with much by way of reliable insight into abstract metaphysical matters – matters far removed from the exigencies of normal human life? Surely the answer here is "no." According to the usual evolutionary story, human cognitive faculties – of the same basic sort we possess today – appeared during the Pleistocene era, during the period lasting from about 2.5 million years ago to about 12 000 years ago. By the end of that period, our ancestors possessed brains of roughly the same basic architecture and cognitive capabilities as our brains today. The main explanation for why they evolved those faculties, on the usual story, is that having such faculties was adaptive in Pleistocene environments: useful for feeding, flying, fighting, and reproducing on the plains of Pleistocene Africa. But why would cognitive faculties selected for success at those tasks in those environments have required reliability with respect to abstract metaphysics, reliability with respect to questions about the deep structure of realities underlying the appearances, knowledge of which would've been wholly irrelevant to life on the plains of Pleistocene Africa? From a fitness point of view, such cognitive capability seems wholly unnecessary. But if so, the probability that it should have evolved seems low.

There is, however, the possibility that reliability on abstract metaphysical matters far removed from the everyday concerns of life is a "spandrel" – a nonadaptive byproduct of some adaptively selected trait, just as, for example, reliability on abstract mathematical matters is plausibly thought of as a nonadaptive byproduct of the adaptive ability to do simple arithmetic. That could be, but as I read the cognitive science literature, no one has been able to find good reason for thinking so. More, as I'll try to argue further on, the prospects for producing a decent argument for thinking so are dim. The right thing to say, I think, given the current state of the evidence and the considerations I raise further on, is that the probability we should have got cognitive faculties capable of reliability on abstract metaphysical matters, given naturalism and the usual evolutionary story, is inscrutable (we have no way of knowing what it is).

Let me pause briefly to note that I am not here recapitulating Alvin Plantinga's famed "evolutionary argument against naturalism,"[3] in which Plantinga argues that the probability of our overall cognitive reliability (the probability that our faculties should be generally truth-conducive, with respect even to everyday matters), given naturalism and evolution, is low or inscrutable, and then goes on to argue that this gives the believer in naturalistic evolution reason to doubt all of the deliverances of her cognitive faculties, including belief in naturalistic evolution. My argument is inspired by Plantinga's argument and similar in many ways,[4] but it differs in this key respect: I am not claiming that, given naturalistic evolution, the probability that our cognitive faculties are generally reliable is low or inscrutable. I am arguing for something much weaker. Objectors to Plantinga's argument have claimed that, contrary to Plantinga, given naturalistic evolution, we should expect our faculties to be generally reliable.[5] I won't dispute that here. I merely claim that, given naturalistic evolution, the probability that our cognitive faculties are reliable *with respect to*

3 See, e.g., Plantinga (2000) and Beilby (2002).

4 My argument also bears close affinities to arguments developed in two earlier papers of mine (Crisp 2011; 2013), though below I'll develop a couple new objections to the argument of those papers. Argument in the same basic family may also be found in Otte (unpublished), Rea (2002), Reppert (2003), and Nagel (2012).

5 See, e.g., Fitelson and Sober (1998), Fales (2002), Fodor (2002), and Ramsey (2002).

abstruse metaphysical matters is inscrutable. And that might well be, even if, given naturalism and evolution, we should expect some sort of general cognitive reliability.

Finally, there is this:

(3) *Principle of Reason:* If, for some source of information S, you have good reason to doubt the reliability of S, and have no good reason to discount this reason for doubt, then the rational attitude toward matters about which S is your only source of information is doubt.[6]

The key terms here may be understood as follows. You have *good reason to doubt* the reliability of a source of information S if there is some condition C_1 such that you have good reason to think that S is in condition C_1 and good reason to be agnostic about the probability that S is reliable given that it is in C_1. You have *good reason to discount this reason for doubt* if there is some further condition C_2 such that you have good reason to think that S is in both C_1 and C_2 and good reason to think that S is probably reliable given that it is in both conditions C_1 and C_2. Finally, the rational attitude to take toward some matter is *doubt* if the rational attitude, given your evidence on the matter, is either agnosticism or disbelief.[7]

Why accept the Principle of Reason? I argue by cases. Suppose you have good reasons for agnosticism about the accuracy of a certain pool thermometer (say, you stepped on it and the gauge now rubs against the metal backing), no reason for thinking that, in this case, those reasons don't apply, and no further information about the water temperature in the pool. Then, pretty clearly, you shouldn't believe what the thermometer says about the water temperature; the rational response here is agnosticism. Or suppose you have good reason for agnosticism about Uncle Fred's reliability on matters of family history (he is suffering from mild dementia), no special reason for thinking that, in this case, he can be trusted, and no further information about the stories he is now relating. The rational response here, one thinks, is doubt. Reflection on these and like cases, I suggest, lends strong support to our principle.

These theses and principle in hand, I can now state my objection to naturalistic metaphysics. Suppose you are a practitioner of abstract naturalistic metaphysics, proposing to explain the mental, the moral, and the physical in purely naturalistic terms, and along the way postulating structures and objects quite different than those encountered in everyday experience, connected in highly complex ways to the appearances, knowledge of which is far removed from the everyday concerns of life and subject to perennial dispute. Then, I suggest, you have good reasons for accepting both of our theses – the Evolutionary Thesis and Thesis of Unreliability – as I have already laid out. But if so, you have good reason to doubt the reliability of those parts of your cognitive endowment responsible for abstract metaphysical beliefs (henceforth, your *metaphysical faculties*). This is because you have good reason to think that your metaphysical faculties are in the condition one's faculties would be in if they were produced by unguided processes of the sort described by contemporary evolutionary theory, and good reason for agnosticism about the proposition

6 Cf. Rea (2002, 186ff.), who deploys a similar principle, in argument similar to the one I'm developing here, against certain forms of intuitionism: the thesis that rational intuitions and the methods of natural science are the only basic sources of evidence.

7 I shall understand talk of *disbelief* here and in the sequel in such a way that someone disbelieves a proposition p if and only if she either fully believes not-p or invests a considerable degree of confidence in not-p. I understand talk of *agnosticism* here and in what follows in such a way that agnosticism regarding a proposition p is incompatible with considerable confidence in either p or not-p.

that one's metaphysical faculties are probably reliable given that they are in that condition. Assuming you have no good reason to discount those reasons for doubt – no good reason for thinking your metaphysical faculties are in some further condition such that they are likely to be reliable – then by our Principle of Reason, the rational attitude for you to take toward the deliverances of your metaphysical faculties is doubt: agnosticism or disbelief. Assuming you have no good reason, as we'll put it, to doubt the doubts I have raised, then, if you are a practitioner of abstract naturalistic metaphysics, the rational course for you is systematic doubt toward the results of your work. And that's a deep problem for naturalistic metaphysics, since presumably the point of the enterprise is to arrive at claims one can believe, or at any rate accept with considerable confidence. It's a problem, I'd think, for the authors of Chapter 2, since, one suspects, their attitude toward their various claims about metaphysical freedom, agent causation, the nonexistence of a faculty of the will, there being no irreducible, non-natural moral properties, all moral truths being grounded in facts about human needs, desires, and purposes, and so forth, is neither agnosticism nor disbelief. If they have no good reason to doubt these doubts, however, then the rational response to these various claims of theirs is either agnosticism or disbelief.

The big question, then: Do practitioners of naturalistic metaphysics have good reason to doubt the doubts I have raised about the reliability of those parts of our cognitive endowment responsible for abstract metaphysical beliefs? I don't think so, but let us look into some possibilities.

Reasons to Doubt the Doubts?

The "Miracle" Argument

There is an argument in the philosophy of science literature for scientific realism – the view that our scientific theories are true or approximately true, as opposed to just useful predictive devices – that might be useful here: the so-called "miracle" argument. It takes its start from the observation that our best scientific theories are extraordinarily successful: they enable remarkably accurate predictions of a huge variety of phenomena, and enable us to manipulate and control our environment in myriad ways. The argument suggests next that by far the best explanation of this is that our theories are true, or at least approximately true; for if they weren't, it would be nothing short of a miracle that they have been so successful. The most reasonable thing to think, the argument concludes, is that our theories are true or approximately true.

This has obvious application to my claim that we have no reason to doubt the doubts I raise about the reliability of our metaphysical faculties. For, if the miracle argument gives good reason for thinking our scientific theories true or approximately true, then you might think it also gives good reason for thinking our metaphysical faculties reliable. Our best scientific theories involve quite a lot of abstract metaphysical theorizing (think relativity and quantum physics), and, if they are true or near true, then our metaphysical faculties would seem to be reliable.

To make the same point differently, suppose our metaphysical faculties are unreliable. How to explain, then, our remarkable success at predicting and controlling the huge variety of phenomena we manage to predict and control by way of scientific theorizing? By far the best explanation of this, you might argue, is that our metaphysical faculties are reliable; for if they weren't, it'd be nothing short of miraculous that our scientific theorizing has been so

successful. So the reasonable thing to think here, you might conclude, is that our metaphysical faculties are reliable and that we *do* have good reason to doubt the doubts I raise about their reliability.

Note by way of reply, though, that there are other ways of explaining the success of scientific theories, ways incompatible with scientific realism, such as Bas van Fraassen's constructive empiricism. On this view, theory change in science is a sort of evolutionary process (van Fraassen 1980, 39ff.), a process that selects for theories that are *empirically adequate* (accurate in predicting and describing the observable world), weeds out theories that aren't empirically adequate, but is blind to what we might call *metaphysical adequacy*: accuracy in mirroring the unobservable structure of things. It's no miracle, says van Fraassen, that an evolutionary process of this sort should produce successful (i.e., empirically adequately) but false (i.e., metaphysically inadequate) theories: you'd expect as much from such a process.

Now, it's no part of my project to argue that van Fraassen's constructive empiricism is correct. I bring it up only to point out that you'd have good reason for thinking the reliability of our metaphysical faculties the best explanation of our theoretical successes only if you had good reason for ruling out alternative explanations of the relevant successes, like van Fraassen's. The trouble is, there is furious debate in the philosophy of science over these matters, and the issues relevant to deciding them are quite, well, abstract. Deciding whether scientific realism or an alternative (nonrealistic) explanation is the best explanation of our theoretical success, then, is itself an exercise in abstract metaphysics. Since deciding between these competing explanations is a crucial step in the miracle argument, we see that the miracle argument itself is a fine example of abstract metaphysics.

But no such argument could, all by itself anyway, constitute good reason for thinking your metaphysical faculties reliable. Suppose the defense attorney proposes to convince the jury that so-and-so is a reliable witness. The attorney's only evidence: the witness in question *says* of himself that he is reliable. That's pretty weak evidence; that a witness assures us he is a reliable witness isn't much reason for thinking so. Likewise here: that you have an abstract metaphysical argument whose conclusion is that your metaphysical faculties are reliable isn't much reason for thinking so. Such an argument presupposes the reliability of the very faculties it purports to show the reliability of.[8] You couldn't sensibly accept the premises and method of reasoning in such an argument unless you *already* had some justification for thinking your metaphysical faculties reliable, justification that was prior to and independent of this argument. If your *only* evidence for thinking your metaphysical faculties reliable were this argument, it would be weak evidence indeed.

I conclude that, if the miracle argument is your only reason for doubting the doubts I raised about the reliability of your metaphysical faculties, then you have no good reason for doubting those doubts.

We Should Expect Metaphysical Reliability from Creatures Capable of Abductive Reasoning

Trent Daugherty offers another reason for doubting those doubts (2012; I'll put the objection in my own way, but the basic idea is his). Consider the fact that we have good reason to think that our Pleistocene forbearers were capable of abductive reasoning: reasoning to

8 Cf. Alston (1991, ch. 3).

hypotheses about unobserved phenomena as the best explanation of observed phenomena. (This comes to us from anthropological studies of primitive hunter-gatherer societies, whose members are quite similar in terms of culture and cognitive power to humans living at the end of the Pleistocene era, and who make use of sophisticated abductive reasoning in hunting and tracking animals; see, for example, Carruthers 2004.) Somehow or other, our ancestors evolved that capacity. But once they had it, it's not surprising that they would have been capable of reliably redeploying it to previously unencountered domains. Think about arithmetic: once you have acquired the ability to reliably count buffalo, it's not terribly surprising that you will then be capable of reliably counting tigers. Likewise with deductive reasoning: once you have acquired the ability to validly infer truths in some one domain, it's no great surprise that you are thereby capable of valid inference in other domains as well.

Likewise with abductive reasoning. Our ancestors evolved faculties capable of reliable inference to the best explanation in the domains they encountered on the plains of Pleistocene Africa. That ability was adaptive and so passed along to their descendants. We have simply reapplied that method of reasoning to other domains, including the domains studied by abstract metaphysics, and it's no more surprising or improbable that it should be reliable in these other domains than it is that arithmetic or deductive reasoning should be reliable across a variety of domains.

We can put the point like this: True enough, we have good reason for thinking our metaphysical faculties the product of naturalistic evolution and good reason for agnosticism about the probability that they should be reliable given that they are the product of naturalistic evolution. But we also have good reason for thinking our ancestors developed the ability to reason abductively and for thinking it likely that creatures capable of abductive reasoning should be capable of reliable metaphysical reasoning. The latter constitutes good reason for doubting the doubts raised about the reliability of our metaphysical faculties.

So says our objector. By way of evaluation, let us look into the nature of abductive reasoning more closely.[9] Abductive reasoning, once again, is reasoning to hypotheses about unobserved phenomena as the best explanation of some observed phenomena. You emerge from a night's sleep to discover a bowl and spoon covered with dried ice cream on the counter. There are a variety of possible explanations: your house was broken into and the intruder enjoyed a bowl of ice cream but left all else untouched; your in-laws visited last night, had a bowl of ice cream, then left; and many more. But the best explanation is that your child had a late night snack before bed. Since that's the best explanation of your observational evidence, that's the one you believe.

Note two crucial features of the abductive process. First, there is a creative component: one imaginatively constructs possible explanations, where often these involve unobserved or unobservable mechanisms. In the reasoning of the previous paragraph, this part goes quickly, and you likely don't do much by way of imaginative construction of possibilities. Given your evidence, you likely need only consider a couple possibilities (or even just one). In other cases – in the sciences, for example – this part of the process is considerably more involved: one spends enormous time and effort imaginatively constructing various explanatory mechanisms. The key point: there is a creative moment in abductive reasoning, involving imaginative construction of explanatory mechanisms.

Second, there is a set of selection principles guiding one's choice of explanations (by which one determines the best explanation). There is debate among philosophers and

9 I was helped here by Peter Carruthers' (2004, 81ff.) discussion of abductive reasoning.

cognitive scientists about just what this set contains, but a typical list would include features like simplicity (the explanation is capable of simple expression, and postulates fewer entities or types of entities than its competitors), fit with background evidence (the explanation fits with other things one knows about the world better than its competitors), explanatory scope (the explanation explains a wider range of data than do its competitors), and fruitfulness (the explanation suggests new avenues of inquiry). The best explanation is the one that bests its competitors with respect to these and like selection principles.

Reflection on these two aspects of abductive reasoning suggests two requirements on successful use of abductive reasoning. First, there is what we might call the creativity requirement: successful abductive reasoning in some domain requires that we be capable of imaginatively constructing explanatory mechanisms of the sort operative in that domain. Conceivably, a domain of inquiry could be so foreign to us, so beyond our ken, that construction of explanatory mechanisms of the sort operative in that domain would be beyond our imaginative powers. Successful abductive reasoning in a domain requires that that not be the case.

And second, there is what we might call the selection requirement: successful use of abductive reasoning in some domain requires that the selection principles guiding our adjudication between competing theories (simplicity, fruitfulness, etc.) are good indicators of the sorts of explanatory mechanism likely to be operative in that domain. Conceivably, a domain of inquiry could be such that our usual principles of theory selection aren't good indicators of the sorts of mechanism operative in that domain. Successful abductive reasoning in some domain requires that that not be the case.

These reflections on abductive reasoning in hand, let us return to this claim:

> Our ancestors evolved faculties capable of reliable inference to the best explanation in the domain they encountered on the plains of Pleistocene Africa. That ability was adaptive and so passed along to descendants. We have reapplied that method of reasoning to other domains, and it's no more surprising or improbable that it should be reliable in these other domains, including the domains studied by abstract metaphysics, than it is that arithmetic or deductive reasoning should be reliable across a variety of domains.

We can now see that whether it is surprising or improbable that the abductive abilities we inherited from our Pleistocene ancestors are reliably redeployable in other domains, including those studied by abstract metaphysics, turns on whether it is surprising or improbable that our abductive powers satisfy the creativity and selection requirements vis-à-vis those domains. The crucial questions, then: How probable is it that the imaginative powers we inherited from our Pleistocene ancestors would have rendered us capable of imagining explanatory mechanisms of the sorts operative in the domains studied by abstract metaphysics (e.g., the quantum realm, or the realm of abstract objects)? And: How probable is it that the selection principles inherited from our Pleistocene ancestors[10] would reliably discriminate between explanatory mechanisms likely to be operative in the domains studied by abstract metaphysics and those not likely to be operative in those domains?

By way of initial answer to these questions, I'd say: No one knows; no one is in a position to answer these questions. The main explanation why our Pleistocene ancestors evolved the

10 For argument that those principles *were* inherited, see Carruthers (2004, 92ff.).

imaginative powers and abductive principles they did was that having those powers, and making use of those principles, was adaptive in Pleistocene environments: useful for feeding, flying, fighting, and reproducing on the plains of Pleistocene Africa. Now, *maybe* imaginative powers and abductive principles selected for their success at those tasks in those environments would be capable of redeployment to the sorts of domain studied by abstract metaphysics, but it's hardly clear that we should *expect* this. Just as plausibly, the imaginative powers and abductive principles evolved by our Pleistocene ancestors were adaptive vis-à-vis evolutionary problems faced by hunter-gatherers on the plains of Africa but utterly unsuited to the domains studied by abstract metaphysics. As best I can tell, the probabilities in question are inscrutable: no one knows or could know what they are. We don't know how probable it is that abductive abilities evolved by our Pleistocene ancestors should satisfy the creativity and selection requirements vis-à-vis domains studied by abstract metaphysics: the quantum domain, and so on. And if so, though it might be right that we have good reason for thinking our ancestors evolved the ability to reason abductively, we don't have good reason for thinking it likely that creatures with such cognitive capability should also be capable of reliable metaphysical reasoning. We don't, so far, have good reason for doubting the doubts I have been on about.

A line of objection worth briefly exploring goes like this: "You are presupposing that the imaginative powers and abductive principles we modern humans deploy in our metaphysical theorizing are the same powers and principles our Pleistocene ancestors deployed in their theorizing. But this assumption is unwarranted. There is every reason to believe that these powers have evolved over the thousands of generations that separate us from our Pleistocene forbearers. And if so, it's no reason to doubt the reliability of *our* abductive practice in abstract metaphysical domains that *their* powers of imagination and abductive selection principles may well have been inadequate to metaphysical theorizing."

There is a debate in cognitive science that is relevant here. Both sides of the debate accept a computational view of the mind, according to which the brain is an extremely complex digital computer and thought is a kind of computation. Both sides hold that the *hardware* of the brains of our Pleistocene forbearers is essentially the same as the hardware of our modern brains. The innate structure and consequent computational capacities of their brains differed very little from the innate structure and consequent computational capacities of our brains. (There was too little time between them and us for substantial hardware evolution.)

There is disagreement, however, about *software* differences between us and our forbearers.[11] One side of the debate holds that the software running on our brains has been thoroughly reprogrammed since the time of our Pleistocene forbearers. According to Daniel Dennett (1991), for example, software is encoded in human brains by language use and imitation, especially in childhood. The contents and information-processing algorithms encoded in this software are determined by a person's cultural environment, so that as culture evolves over time (new things learned about how the world works, new methods of reasoning discovered), so too do the contents and information-processing patterns programmed into young human minds. Cultural evolution, then, brings in its train software evolution, with the result that our brains are running software of an extremely different sort than that of our distant, Pleistocene ancestors.

11 My discussion here is heavily indebted to Carruthers (2004).

Another side of the debate holds that there are relatively few software differences between us and our Pleistocene ancestors.[12] Their software enabled them to perform a host of information-processing tasks that were useful in solving the various evolutionary problems they encountered. We inherited from them the same basic suite of software, with the same basic information-processing strategies. Proponents of this "continuity view"[13] of our mental software differ over the nature of the information-processing strategies we inherited. Some think our information-processing abilities "domain-specific," so that our cognitive system is a collection of specialized faculties responsible for producing belief about particular domains. For example, some think we have a specialized belief-producing faculty responsible for producing beliefs about the mental states of others on the basis of observations of their face or body language; a distinct, specialized belief-producing faculty responsible for producing beliefs about the behavior of objects in motion; a further distinct faculty for beliefs about which objects are living and which are mere artifacts; and many more (e.g., Pinker 1997). Others hold that our information-processing abilities are "domain-general," so that we have a collection of belief-forming strategies that apply generally, across the variety of domains we encounter, such as deductive logic and the logic of probability (e.g., Gopnik and Melzoff 1997). But the key point about the continuity view, for our purposes, is that there are relatively few software differences between us and our Pleistocene ancestors: their software enabled them to perform a host of information-processing tasks that were useful in solving the various evolutionary problems they encountered, and we inherited from them that same basic package of software with those same basic information-processing strategies.

The continuity/no-continuity debate bears on our objection in obvious ways. If the continuity side is right, there likely isn't much difference between the imaginative powers and abductive principles deployed by our Pleistocene ancestors and those deployed today, and the points I raise about our being properly agnostic about the adequacy of those imaginative powers and abductive principles to abstract metaphysics goes through as before.

Suppose the no-continuity side is right, though. What then? Might it be that the imaginative powers and abductive principles developed in us since our Pleistocene days have developed in the direction of metaphysical reliability?

Of course it might, but I think we have no way of knowing: that we have no way of knowing whether the software upgrades from our Pleistocene predecessors conduced to abductive reliability in metaphysical domains. How could we know? It's not self-evident or by any means obvious that the software upgrades delivered to us by cultural evolution should have conduced to reliability in metaphysical domains. Perhaps the brain hardware evolved by our Pleistocene ancestors cannot be programmed by language-based instruction and imitation so as to subtend the imaginative powers and abductive principles requisite for reliability in the quantum domain and other domains of abstruse metaphysical inquiry. Chimpanzee brains can't be so programmed, for example, or so it would seem: no regime of language-based instruction or imitation can encode software in their brains sufficient to make them good at metaphysical theorizing. Perhaps the hardware of our brains is similarly constrained. It looks to me like the only reason one could have for thinking the hardware of our brains not so constrained, and for thinking that the software upgrades we inherited

12 So says the "evolutionary psychology" approach of Leo Cosmides and John Tooby (see, e.g., Barkow, Cosmides, and Tooby 1992), Steven Pinker (1997), and collaborators.

13 I borrow this terminology from Carruthers (2004).

tended in the direction of metaphysical reliability, is some sort of abductive argument: argument, first, that the no-continuity view is the best explanation of this or that range of phenomena and, second, that the conjunction of the no-continuity view and the claim that the brain hardware evolved by our Pleistocene forbearers underwent software upgrades of the sort requisite for metaphysical reliability is the best explanation of some further range of phenomena.

But any such argument would itself constitute an exercise in abstract metaphysics. It'd involve quite a lot of postulation of structure far removed from the world of everyday appearance, related to the world of appearances in highly complex ways, and so forth. And for reasons we saw earlier, that you have an abstract metaphysical abduction whose conclusion is that your abductive faculties are reliable in metaphysical domains isn't much of a reason for thinking them reliable. Such an argument presupposes the reliability of the very faculties it purports to show the reliability of. You couldn't sensibly accept the premises and method of reasoning in such an argument unless you *already* had some justification for thinking your abductive faculties reliable in metaphysical domains, justification that was prior to and independent of this argument. It's hard to see, though, what sort of justification that could be.

I conclude that, even if it's right that our imaginative powers and abductive principles have changed over time, we have no good reason for thinking they changed in the direction of metaphysical reliability, and thus no good reason here for doubting my doubts about the reliability of our metaphysical faculties.[14]

Help from Thomas Reid

The 18th-century Scottish philosopher Thomas Reid famously argued that belief in the reliability of our cognitive faculties is not plausibly thought of as based on argument or inference from other beliefs we hold. According to Reid,[15] belief in the reliability of our faculties is based, rather, on a certain sort of *experiential* evidence, viz. an experience of the emotion of *ridicule*. He thought of belief in the basic truths of logic, arithmetic, and "common sense" as grounded in this emotion, as follows: When one considers the contrary of such truths – when one entertains the possibility, for example, that 1 + 1 is *not* equal to 2 – an emotion of ridicule naturally arises, prompting in one thoughts like "that's absurd, ridiculous!" Such experiences, thought Reid, ground – make reasonable – belief in those truths of common sense whose contraries present themselves to one as thus ridiculous. In the language I've been using: such experiences constitute good reason for accepting such truths.

Perhaps we have such reason to doubt the doubts I raised earlier about the reliability of our metaphysical faculties. Perhaps belief in their reliability is like belief in the truths of common sense discussed by Reid: such that, when you consider the possibility that the contrary is true, that your metaphysical faculties are *not* reliable, you find yourself in the grip of a powerful emotion of ridicule. If so, we could argue à la Reid that you have good reason for

14 Similar considerations apply to the possibility, considered earlier, that reliability on abstract metaphysical matters is a "spandrel" – a nonadaptive byproduct of an adaptively selected trait. I said when we first discussed it that the prospects for producing an argument that reliability on abstract metaphysics is a spandrel are dim. We can now see why. Any such argument would constitute both a piece of abstract metaphysics and an argument for the reliability of our metaphysical faculties, and, for the reasons we have been considering, such arguments offer little by way of independent support for their conclusion.

15 See, for example, Reid (1969, 606–607).

thinking those faculties reliable, and good reason for doubting the doubts I raised about their reliability. The reason doesn't take the form of an *argument* that they're reliable: it's a nonpropositional, experiential reason. But no matter: it's a good reason all the same.

By way of reply, though, I wonder this: Why don't *I* have this Reidian experiential evidence for the reliability of *my* metaphysical faculties? I don't experience any emotion of ridicule when I entertain the possibility that my cognitive faculties aren't reliable with respect to abstruse matters of metaphysics far removed from the everyday concerns of life. That possibility doesn't strike me as ridiculous at all. In fact, when I consider the multitude of crazy metaphysical views philosophers have defended over the centuries and the rampant disagreement among theorists over almost everything of substance on questions of metaphysics, I find it wholly *unobvious* that we humans, myself included, have reliable metaphysical faculties. More, that attitude strikes me as *appropriate*, as the right response to the evidence. Given the history of crazy views defended over the years and rampant disagreement among practitioners of metaphysics, one *shouldn't* find ridiculous the possibility that our metaphysical faculties are not reliable – one shouldn't find it just *obvious* that our metaphysical faculties are reliable. Those who do find it obvious aren't, I think, being appropriately responsive to the evidence. But if so, then if your only evidence for belief in the reliability of your metaphysical faculties is Reidian experiential evidence, you *don't* have good reason to doubt the doubts I raised about their reliability.

Conclusion

I have argued that commitment to naturalism gives good reason to doubt the reliability of those cognitive faculties responsible for abstract metaphysical theorizing. I claimed that, if there aren't good reasons to doubt these doubts, then commitment to naturalism gives one good reason for agnosticism about the deliverances of those faculties. I canvassed some main reasons for doubting these doubts. None was impressive. Since I think there aren't any better reasons, I conclude that there is trouble for the project of attempting to explain the mental, the moral, and the physical in naturalistic terms. Any such project will be an exercise in abstract metaphysics, suggesting that, if you are a participant in that project, the rational course for you is systematic doubt toward the results of your work.

References

Alston, W.P. (1991). *Perceiving God: The Epistemology of Religious Experience*. Ithaca: Cornell University Press.

Barkow, J., Cosmides, L., and Tooby, J. (1992). *The Adapted Mind*. Oxford: Oxford University Press.

Beilby, J.K. (2002). *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*. Ithaca: Cornell University Press.

Carruthers, P. (2004). The Roots of Scientific Reasoning: Infancy, Modularity and the Art of Tracking. In *The Cognitive Basis of Science*, edited by Peter Carruthers, Stephen Stich, and Michael Siegal. Cambridge: Cambridge University Press.

Crisp, T.M. (2011). An Evolutionary Objection to the Argument from Evil. In *Evidence and Religious Belief*, edited by Kelly James Clark and Raymond Van Arragon. Oxford: Oxford University Press.

Crisp, T.M. (2013). On Coercion, Love, and Horrors. *Religious Studies* 49: 165–179.

Dawkins, R. (1986). *The Blind Watchmaker*. New York: W.W. Norton.

Dennett, D. (1991). *Consciousness Explained*. London: Penguin.

Dougherty, T. (2012). Review of *Evidence and Religious Belief*. *Notre Dame Philosophical Reviews*. Available from: http://ndpr.nd.edu/news/28918-evidence-and-religious-belief (last accessed July 15, 2015).

Fales, E. (2002). Darwin's Doubt, Calvin's Calvary. In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by James K. Beilby. Ithaca: Cornell University Press.

Fitelson, B. and Sober, E. (1998). Plantinga's Probability Arguments against Evolutionary Naturalism. *Pacific Philosophical Quarterly* 79(2): 115–129.

Fodor, J. (2002). Is Science Biologically Possible? In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by James K. Beilby. Ithaca: Cornell University Press.

Gopnik, A. and Melzoff, A. (1997). *Words, Thoughts and Theories*. Cambridge, MA: MIT Press.

Nagel, T. (2012). *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. Oxford: Oxford University Press.

Otte, R. (Unpublished). Science, Naturalism, and Self-Defeat.

Pinker, S. (1997). *How the Mind Works*. London: Allen Lane.

Plantinga, A. (2002). Introduction. In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by James K. Beilby. Ithaca: Cornell University Press.

Ramsey, W. (2002). Naturalism Defended. In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by James K. Beilby. Ithaca: Cornell University Press.

Reid, T. (1969). *Essays on the Intellectual Powers of Man*. Cambridge, MA: MIT Press.

Reppert, V. (2003). *C.S. Lewis's Dangerous Idea: In Defense of the Argument from Reason*. Downer's Grove: InterVarsity Press.

van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

van Inwagen, P. (2009). *Metaphysics*. Boulder: Westview Press.

6

Naturalism and Realism in the Philosophy Science

MATTEO MORGANTI

Introduction

It is not among the objectives of this essay to offer a precise definition of naturalism – those looking for one can find it in other chapters of this collection. For present purposes, it is sufficient to understand naturalism in broad terms, as the requirement that philosophy be in some sort of "contact" or "continuity" with science. This entails that philosophical investigation should be concerned with, and grounded in, the world as it is known *a posteriori* via our experience and, in particular, through scientific (i.e., observation- and experiment-based) inquiry. In this broad sense, naturalists believe that no attempt should be made at doing so-called "first philosophy," the enterprise whereby big, fundamental questions are dealt with in a completely *a priori* fashion. The obvious rejoinder is that science includes disciplines, such as mathematics, that do not have to do with material things, and consequently appear to transcend experience and indeed proceed purely *a priori*. Setting aside the question of how naturalists should look at mathematics and abstract entities in general, however, here I will focus on the natural sciences only, assuming that a naturalistic requirement formulated along the lines just suggested is clear and compelling, at least when restricted in this fashion.

Since the naturalistic stance explicitly takes science and scientific knowledge to be fundamental, it is obvious that the role played by naturalistic or antinaturalistic assumptions in philosophy should be studied with particular care when it comes to the proper part of the discipline that deals directly with science itself; that is, the philosophy of science. This is so, first of all, because the philosophy of science is the branch of philosophy that lies closest to what naturalists regard as a model for philosophical inquiry and, second, because, in this particular case, it is the philosophical analysis of science itself that the naturalistic viewpoint requires to become "scientific." The foregoing means that naturalists should have something to say about (1) how more or less traditional philosophical questions concerning science should be framed and dealt with by naturalists; but also about (2) how science

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

should be conceived of and studied in a naturalistic context. There is, in particular, one important area where these sets of issues converge and overlap: the debate between so-called scientific realism and antirealism. Indeed, the opposition between these two views on the epistemological import of scientific theories is one of the most fundamental themes (if not *the* fundamental theme) in the philosophy of science. Do our best scientific theories tell us anything true (or approximately true) about the world in which we live, as scientific realists contend? Should we believe, for instance, that Higgs bosons *really* exist out there independently of us, or that the history of the universe *really* began with a Big Bang around 13 billion years ago? Or are scientific theories best conceived of as tools for a successful interaction with the world itself, to be evaluated in terms of practical aims and results rather than knowledge of an objective reality, as antirealists have it – in which case, believing in things like Higgs bosons and the Big Bang is only instrumental to "saving the phenomena"; that is, giving an account of the observable facts around us? And how should we explain, on each of these alternative accounts, the fact that often we can derive surprising predictions from our theories, and devise ways of testing them empirically? These questions can be translated directly into questions with a glaring naturalistic flavor: How is the scientific enterprise to be evaluated from the point of view of our cognitive abilities and their being shaped by natural evolution? Is there any reason to expect evolution, to the extent that it includes among its effects the progress of our scientific knowledge and capabilities, us closer to the truth? But also: Is scientific realism itself a scientific hypothesis, in the sense that empirical data can be brought to bear on its critical evaluation?

These are the interrogatives on which we will focus in what follows, together with another important, corollary one: What does our best current science say about the world around us? The question here concerns *interpretation*; that is, what description of their domain of application we should extract from the theories that scientists regard as the best theories of those domains. On the one hand, this is a corollary question because it is in principle avoidable. Those not endorsing scientific realism, it would seem, can coherently sidestep interpretational issues. And, as a matter of fact, many – perhaps most – contemporary realists are not interested in answering "What is it like?" queries either, and consequently refrain from choosing an interpretation of, say, quantum mechanics or evolutionary biology. Indeed, it is not contradictory to believe that a theory is (approximately) true and/or that the development of science approaches truth without also endorsing any particular interpretation of any specific theory. (In this sense, the systematic avoidance of interpretational issues surely makes the realist position logically stronger because less committing.) On the other hand, the question of interpretation is important nonetheless, and there are other philosophers who explicitly include the task of interpreting our best scientific theories in the realist agenda (I will not discuss here whether questions of interpretation can be meaningfully posed in an antirealist context), thus subscribing to a more ambitious program. Such a program would seem to include more than "just" science, and to require one to say something about the relationship between science and *metaphysics*. For, as we will see in more detail later, at least some genuine metaphysics would seem to be involved in the interpretation of scientific theories. What can be said about all this from the naturalistic viewpoint? What space does proper naturalistic methodology leave for the interpretation of scientific theories? And what should naturalists say about the relationship between science and metaphysics?

In what follows, I will first deal with naturalism in relation to the scientific realism issue. Then, I will more or less assume scientific realism, and discuss the question concerning the

relationship between science and metaphysics in a naturalistic context. In both cases, a range of options and alternatives will be considered, and some arguments additionally offered in favor of a "mild" or "liberal" naturalism in the philosophy of science. The upshot will be that it does indeed make sense both to locate the scientific realism issue in a naturalistic perspective and to demand a naturalistic approach to metaphysics – but only to the extent that this is not taken to mean that one should proceed *exclusively* by empirical means, that scientific realism is a *fully* empirical hypothesis, and/or that metaphysics should either be eliminated or somehow "extracted" directly from science. To the contrary, I will argue that – unless it is given a strict formulation, which, however, sacrifices some interesting and by no means impracticable areas of inquiry – naturalism is best understood in the sense that there are some aspects in which philosophy should be made responsive to science; but, at the same time, the former has to remain different from the latter in essential respects, and should consequently be regarded as autonomous.

Naturalism and Scientific Realism

Naturalism undoubtedly has an empiricist component. In the period from the second half of the 16th to the first half of the 17th centuries, Francis Bacon criticized the Scholastics and, more generally, the Aristotelian tradition for being unable to engage in a direct investigation of nature; and, in the 17th–18th centuries, John Locke and David Hume took it to be obvious that it is impossible to go beyond experience, making this fundamental assumption the cornerstone of their philosophical "systems." Similar ideas were expressed by the thinkers associated with what is known as "Enlightenment" around the 18th century and by the positivists in the 19th century, who also emphasized the privileged epistemic status and social role of science. These traditions were put together, elaborated upon, and turned into a systematic, large-scale program by the neopositivists, or logical empiricists, of the Vienna Circle. In spite of the fact that the neopositivist ideas concerning verifiability, confirmation, the unification of science, and the reduction to sense data have been questioned, and the overall program ultimately superseded, positivism's underlying "philosophical spirit" persists. The paradigmatic contemporary empiricist is no doubt Bas van Fraassen (e.g., 1980; 2002), some of whose ideas we will discuss later. Here, though, we are more generally interested in contemporary expressions of the basic methodological idea according to which philosophy should consist in nothing more than the systematization of science via formal tools and conceptual analysis, and the "purification" of our knowledge, language, and forms of inquiry from everything that lacks an empirical basis.

Empiricism, Naturalism, and Scientific Realism

One interesting contemporary example of this idea at work comes from Jack Ritchie (2008), who defends a sort of "deflationary naturalism" that – much in line with the tenets of logical empiricism – urges philosophers to limit themselves to scientific claims and hypotheses as much as possible. According to Ritchie, the general principle that the naturalistic philosopher has to follow is to pursue only those philosophical projects that can be carried out through a detailed investigation of science. As an example, Ritchie provides the study of the mutual relationships between the different sciences. Philosophers have often been concerned with such connections, in some cases promoting reduction and unification, in others endorsing

a view of science as "patchwork." However, Ritchie argues, the direct observation and study of actual science shows that neither of these two extreme views is correct, and any further reflection must engage with science much more directly and in much more depth than has been done so far. (According to Ritchie, this is likely to lead to a flexible perspective, able to integrate aspects of the reductionist and antireductionist stances within a diversified picture, in which real-world differences between distinct domains, theories, and scientific disciplines are satisfactorily taken into account).

Independent of the specific example used by Ritchie, this gives us an initial idea of what role philosophy of science might play in a naturalistic context. Interestingly for our present purposes, Ritchie takes his viewpoint to be analogous to Fine's (1984) "natural ontological attitude": the view that science should be taken at face value and no philosophical gloss should be added to it. In particular, Fine and, following him, Ritchie opt for a neutral stance toward the issue concerning the epistemic value of scientific theories, based on the idea that no empirical support can ever be found for either scientific realism or scientific antirealism. At most, they argue, we might take note of the "light" realist attitude exhibited by practicing scientists, whereby – to put it in Carnapian terms – claims such as "gravitons exist," "the molecular unit of heredity for a living organism is the gene," and "the mass of the Higgs boson is between 124.8 and 126.4 GeV/c^2" are only meaningful "internally." That is, they are meaningful only within the scientific discourse, and should not be interpreted as claims about what *really* exists out there (we will come back to the explanatory power of claims concerning truth and existence later). Analogous views can be found in other contemporary philosophers. For instance, Penelope Maddy promotes what she calls "second philosophy," a way of doing philosophy that she regards as a compelling, austere form of naturalism. Essentially, second philosophy ignores big philosophical questions and systems, and recommends instead the use, and reference to the results, of "what we typically describe with our rough and ready term 'scientific method'" (Maddy 2007, 2).

In general, "radical naturalists" such as Ritchie and Maddy share with empiricists of various epochs the belief that verifiable experience (certified by science) is all that counts, and that we should systematically "minimize epistemic risk" by refusing to go beyond that. And such a belief does appear to lead, more or less directly, to either antirealism or the sort of deflationist/agnostic stance described a moment ago. However, it is not obvious that there is this tight connection between empiricism and naturalism, or at least that such a connection cannot be significantly loosened. Perhaps it is possible to be a naturalist and, at the same time, to subscribe to some form of scientific realism? To say more on this, it is necessary to take into account other aspects of the notions of "naturalism" and "naturalized philosophy of science."

A first thing to consider is that, in more or less recent decades, philosophy, and the philosophy of science in particular, has become more and more naturalistic, not only in what one could call an "object-oriented" sense, but also in a "subject-oriented" sense (this distinction takes its cue from Price's (2004) distinction between "object naturalism" and "subject naturalism"). In the first sense, naturalism is indeed an essentially empiricist thesis about what exists and what we can and should be committed to. Since she gives a central role to empirical data and testability, the object-oriented naturalist cannot but take the priority and authority of science to go together with skepticism toward the unobservable entities and processes that scientific realists and metaphysicians care about (there is an interesting connection here – which we have no space to discuss – with traditional debates about physicalism). In the second sense, however (i.e., as subject-oriented naturalism),

the naturalistic stance gives us rather different indications. On this alternative construal, naturalism is an epistemological, not an ontological thesis. It holds that, since human beings are a part of nature, so too are their creations and activities, including science itself. Consequently, science itself should be studied naturalistically; that is, by looking at the empirical features of scientific theories and practices, as well as of scientists. In this context, crucially, no *a priori* distinction between "good" and "bad" beliefs about what exists needs to be made: this sort of issue simply falls outside of the (primary) scope of the view. That subject-oriented naturalism brings with it the need for a more open-minded attitude than object-oriented naturalism is what, for instance, is explicitly suggested by Price. He claims that – even assuming that philosophy and science "are not simply disjoint, and…philosophy properly defers to science, where the concerns of the two disciplines coincide" (Price 2004, 71) – an inquiry based on a naturalistic methodology may lead us to the conclusion that there is more than science, as not everything falls within its scope. *A fortiori*, ontological commitments (i.e., beliefs about what exists or does not exist) need not be determined exclusively on the basis of science – or, as the point is maybe better put, on the basis of a strictly scientific methodology applied to philosophy.

The key point for our discussion is that, though object-oriented naturalism *prima facie* includes and absorbs subject-oriented naturalism (as human beings would seem to be "nothing but" natural objects, albeit of some special sort), there are good reasons for giving priority to subject-oriented naturalism instead. Isn't the study of x as a product of the natural activities of human beings more fundamental than the naturalization, in one sense or another, of some object y on the basis of such study?[1] Following this route leads to rather different philosophical conclusions than those at the basis of what we have called "radical naturalism." Let us, then, look now at what an alternative, less radical take on naturalism about science might look like.

A useful guiding idea here is that science is a byproduct of natural evolution, much in the same sense in which individual human beings and human societies are. Indeed, a study of cognitive, social, and historical aspects has become crucial in several post-neopositivist philosophical approaches to science. From the 1960s onwards, as is well known, studies of the history and sociology of science started to flourish. The Kuhnian emphasis on the context of discovery of scientific theories, for instance, clearly goes in the direction of considering science as a natural entity subject to natural processes, most notably involving the diachronic evolution of social groups. The same basic tendency appears to underlie forms of sociology-based *constructivism* (see, e.g., Bloor's 1976 "strong programme of the sociology of scientific knowledge") and the analogous constructivist views of scientific facts based on the observation and reconstruction of the actual work of scientists in their laboratories (see, e.g., Latour and Woolgar 1979). Generally speaking, these studies have leaned toward scientific antirealism by emphasizing the linguistic and conceptual *incommensurability* between successive scientific *paradigms* and/or the way in which scientific decisions, evaluations, claims, and perhaps even entities are affected by non-epistemic

1 Price argues that object naturalism needs to be "validated," in the sense that it crucially relies on "representationalist" assumptions – concerning the function of human language with respect to the world that is the object of our knowledge – which have to be critically assessed. He further argues that the only means to carry out such an assessment is by first endorsing and implementing subject naturalism. (Price is skeptical about the possibility of a positive outcome of this process of validation, but we need not go as far as to put object naturalism into doubt here).

factors such as utility, social prestige, peer pressure, and the like – so much so that they might even have to be considered human *artifacts*.[2]

Following a different route, some philosophers have developed the naturalistic approach to science with a view to saving some form of scientific *rationality*, but not the notion of truth central for scientific realism. For example, Lakatos (1971) and Laudan (1987) both offered a reconstruction of scientific progress based on the concepts of rational inquiry, research program, and problem-solving ability. And Laudan (1981), at the same time, put forward an explicit argument against realism.

But this is not all. A different strand of research has, in the last three decades or so, focused on the idea that scientific theories and models are ultimately tools for representing the world and achieving certain aims by interacting with it, much in line with the cognitive capacities expressed in our daily lives. Seminal texts in this sense are Giere (1988) and Kitcher (1993). Both authors agree with exponents of the historical and sociological approaches to science that non-epistemic factors play an important role in defining and selecting the models that scientists use to represent the world. However, they make an effort at defining a more comprehensive model of science, in which all the various elements are put together in a reasonable and balanced way. Interestingly enough, this leads both Giere and Kitcher to endorse scientific realism. The evolution of our cognitive capacities and the ability of scientific research to produce ever-improving representations, they argue, can, and in fact did, lead to models representing real, yet unobservable, parts of the world in a progressively more faithful way. In particular, Giere is an *entity realist*, and thinks that, in cases such as the manipulation of protons to inquire into the atomic structure of certain chemical composites or the spraying of electrons to study the features of certain fundamental forces (see Hacking 1983), scientists are best interpreted as employing models that are more than merely "empirically adequate" (van Fraassen's term of art for those theories and models that "save" (i.e., do not go beyond) the observable phenomena) representations of those entities. Kitcher, on the other hand, claims that a fine-grained look at the complex interplay of epistemic and non-epistemic factors that are distinctive of science warrants a realist conception of scientific progress at the level of theories.

It is worth noting that Kitcher, like van Fraassen, employs a very naturalistic, Darwinian analogy in support of his main argument; but they take the analogy to pull in opposite directions. For van Fraassen, "the success of current scientific theories is no miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive – the ones which *in fact* latched on to *actual [observable] regularities in nature*" (1980, 40; italics added). Meanwhile, Kitcher claims that "theories succeed in this way *because they fasten on aspects of reality* [which are neither limited to what is observable, nor modally constrained]" (1993, 156; italics added).

There is a lot more to say on this, of course, and the issue of scientific realism remains wide open. However, the foregoing suggests that a non-strictly empiricist, naturalistic take on the issue is viable; that it has led to interesting developments in recent times; and that it is worth pursuing further. While this is already relevant for our discussion of naturalism in

2 Examples are offered in this context whereby the way in which an experiment should be performed, the interpretation of the data, the theories that should be given priority over their alternatives, and so on are determined by elements that have nothing to do with the way in which the (non-human) world is: for instance, the authority of a project leader, the possibility of funding, the popularity of a research program, and so on.

the philosophy of science, perhaps more can be said by focusing on a related, but different, set of issues: that concerning the status of the scientific realist hypothesis itself.

Naturalism and the Status of the Scientific Realist Hypothesis

As wonderfully illustrated by Hempel, with his example of Ignaz Semmelweis' work in the general hospital of Vienna between 1844 and 1848 (Hempel 1966, ch. 2),[3] there is nothing that qualifies as conclusive evidence – and hence, there are no deductively warranted hypotheses – in science. In the same way in which Semmelweis had to consider various hypotheses before identifying the *best explanation* for certain facts having to do with the mortality of new mothers in the hospital, so science seems to generally rely on "abductive," rather than deductive, reasoning. That is, the available empirical data are never such that the right explanation for the observed phenomena (i.e., the right theory) can be logically inferred from them; instead, various theories are always available, and we are always forced to choose one of them based on its fit with the evidence (plus other, more pragmatic considerations, such as simplicity, unification, and the like). True, alternatives might be unavailable in practice (indeed, scientists find it difficult to come up with even just one sufficiently explanatory theory for any relevant set of phenomena). Yet, logically speaking, the data always "underdetermine" our hypotheses – the concept of "underdetermination" denoting exactly the previously mentioned one-to-many relation between empirical data and possible theories accounting for them. Good science, then, is the sort of activity that – although it never attains certainty – systematically manages to overcome the obstacle represented by the multiplicity of hypotheses available for a given domain of inquiry, and select the best one available. Now, couldn't a naturalist say the same with respect to realism as a philosophical hypothesis? Putnam (1978) famously claimed that scientific realism is the only hypothesis that wouldn't make the – actually observed! – success of science a miracle. Boyd (1996) and others built upon this, arguing in particular that the capability of theories to unify seemingly independent phenomena based on alleged underlying ontologies and to predict unexpected phenomena is a clear sign that there is indeed a success–truth connection when it comes to scientific theories. Note, this is exactly the same reasoning underlying the arguments for the previously mentioned entity realism: how else could we explain, people like Giere and Hacking ask, our repeated (and repeatable) success in interacting with and using certain unobservable entities via complicated apparatuses, if not on the basis of the real existence of those entities? In a nutshell, then, naturalist scientific realists can present their hypothesis concerning the epistemic import of science as itself a well-supported (perhaps "second-order") scientific hypothesis, able to explain facts concerning science.

However, antirealist replies going along the same lines can be found. Van Fraassen, for instance, rejects inference to the best explanation entirely. Even without going this far, empiricists can argue that scientists are best interpreted as accommodating the available data and providing answers to why-questions that do not entail any ontological commitment

3 Semmelweis managed to explain the unusually high mortality of women who had just given birth and were staying in one of two maternity clinics belonging to the hospital. To do so, he formulated and tested several hypotheses based on the available data, discarding all of them as empirically refuted until he ultimately found what appeared to be the right explanation for the phenomenon (doctors did not properly wash their hands before visiting the new mothers).

and/or commitment to the truth or approximate truth of theories. Analogously, constructivists of various stripes can reject realism based on their reconstruction of science as a real-world enterprise that is affected by non-epistemic factors and that has nothing, or only a little, to do with truth. Thus, even if inference to the best explanation were accepted, relativism, constructivism, instrumentalism, and the like (including, though on a slightly different note, agnosticism) might be regarded as the accounts of science that philosophers should favor. It is worth pointing out that in this context, history (i.e., the analysis and reconstruction of the diachronic development of science) plays an important role. Granting for the sake of argument the realist intuition concerning the link between success and truth, Laudan (1981) objected to the ensuing miracle argument by pointing to the fact that, historically, several successful theories have turned out to be false (from the perspective of later science, of course),[4] and the realist assumption should consequently be considered empirically refuted. Here, too, the dispute opens, with different reconstructions of the historical evidence opposing one another, hence different ways of accounting for the input coming from the real world.

I take the foregoing to lend further substance to the adoption of a naturalistic stance in the philosophy of science, and to the idea that the issue of scientific realism versus antirealism falls within the scope, and is in fact best inquired into from the perspective, a naturalized philosophy of science. First of all, scientific knowledge and the related cognitive capacities are in continuity with our everyday knowledge and cognition, so as to recommend a study of science and scientific theories as natural entities subject to natural processes. Additionally, both scientific realism and antirealism (and the agnostic third way traceable, for instance, in the work of Fine, which we mentioned earlier) are hypotheses to be evaluated, much like their scientific counterparts, on the basis of the form of reasoning that led to their formulation, their explanatory power, and their general connection with empirical (including historical) data. Having said that, I would now like to close this section by adding two important, more speculative, claims.

The first concerns the strength of the evidence that can be brought to bear in science and in philosophy, and what this entails with respect to the way in which naturalism should be understood. In science, experiments can be carried out that are directly relevant to an evaluation of a certain hypothesis (though strictly speaking there are no "crucial experiments," that conclusively verify or falsify a theory, and in principle any theory could be preserved in spite of recalcitrant evidence[5]). Think, for instance, of the confirmation of prediction concerning the bending of light in proximity of the sun, or of the existence of the "white spot" that logically follows from Fresnel's theory of light. In each case, the world seems to provide something very close to a yes/no answer to the question(s) that we formulate. But with respect to scientific realism, the evidence that is called into play is much more open to discussion. For example, in the same way in which, say, van Fraassen can insist, in Cartesian fashion, that we shouldn't be committed to something that is not certain, so the realist can insist on claiming that truth is the best explanation of success. In both cases, the lack of a deductive evidence–theory link means that no argument is likely

4 Think, for instance, of the phlogiston theory of combustion, the theory according to which heat corresponds to a *sui generis* entity (so-called "caloric"), the view according to which light is transmitted by an elastic medium (the "ether"), or even Newtonian mechanics as a whole.

5 Readers will see that there is a connection here with the previously mentioned underdetermination of scientific theories by empirical data.

to convince those not already converted (this I take to be the message in recent work aimed at the potential "base rate fallacy" behind the whole scientific realism discussion – see, for instance, Magnus and Callender 2004 and, for a more general overview of the relevant issues, Howson 2000).

This suggests that it might be wrong to take scientific realism to be – as some naturalists would have it[6] – not only a hypothesis about science but also a scientific hypothesis itself. To the contrary, a rather relevant disanalogy seems to emerge between scientific and philosophical hypotheses with respect to their degree of "closeness" – hence, responsiveness – to the empirical (including historical and sociological) input. Indeed, in view of this, the "strict" naturalist may well be considered justified in endorsing a neopositivist stance of sorts, ruling out both scientific realism and antirealism because neither is fully reducible to science, its methods, and its results (both positions would in this case be regarded as "philosophical" hypotheses in a somewhat pejorative sense).

On the other hand, the difference between these two philosophical hypotheses and their scientific counterparts might be *only of degree*. For, first, both scientific hypotheses and philosophical hypotheses only have an abductive (not deductive) basis; and, second, although they are further apart from the world out there than scientific ones, philosophical hypotheses may nevertheless still be responsive to some form of empirical testing, broadly understood. If this were the case, there would seem to be space for an alternative "mild," "moderate," "liberal" (or what have you – I will call it "moderate" from now on) form of naturalism: one that – besides being "subject-oriented" rather than (just) "object-oriented" in the sense previously defined – identifies *some* continuity between science and philosophy, but also concedes some differences; such moderate naturalism attempts to strike a balance between purely philosophical analysis and purely empirical research, without the effective elimination of (a lot of) philosophy. I will elaborate on this in the next section, but the idea is roughly this: while the radical naturalist takes the priority of science to rule out philosophical theses and concepts – such as scientific realism, the existence of unobservable entities, or even, say, that objects are bundles of universals – moderate naturalists need not do so. To the contrary, the latter might regard the belief in universals or quarks and/or the partial truth of a given scientific theory as justified by the development of science itself, and by the growth of our knowledge more generally, even though it transcends the empirical domain.

This leads us to a second point. Even granting that scientific realism and antirealism are based on conflicting assumptions and intuitions that go beyond what is accepted by strict naturalists (i.e., empirical data and their immediate consequences on our hypotheses), more can be said without necessarily trespassing the limits of acceptably naturalistic philosophy. Indeed, this is the key idea underlying the sort of moderate naturalism just conjectured. Let me start from the previously mentioned analogy between van Fraassen's preference for empirical adequacy and the "Cartesian choice" of not committing oneself to anything that is less than certain based on a "risk as little as possible" maxim. If the analogy holds, I wish to contend, then the empiricist position seems to collapse into either full-blown skepticism or realism. Here's the argument: If they don't want to be skeptics, scientific antirealists should either be realists about the contents of perception or "direct realists" (i.e., realists about the objects in the mind-independent, external world that cause the contents of their beliefs). The first option was chosen by logical empiricists, but has never been

6 Recall the Putnam–Boyd line of reasoning mentioned earlier.

turned into a systematic view of the world (for one, how could we explain intersubjectivity and scientific regularities that are independent of what we perceive?). Hence, the second option has to be preferred. Direct realism, however, is based on nothing but a sort of inference to the best explanation from the data of perception to their alleged causes, as those familiar with Hume's *Treatise* know well.[7] And if one can infer from my perceptions and their general features (repeatability, order, explanatory power, etc.) to the existence of something external, *the existence of which I cannot be deductively certain of*, then why exactly should one not do the same with objects of "scientific perception"? As a matter of fact, considerations of continuity seem to suggest two things: first, *pace* detractors of scientific realism as best explanation, truth and existence can indeed play the role of genuine *explanantes*; second, there is no obviously compelling reason to limit our ontological commitments to observable, "medium-sized," "common sense" entities. Thus, if full-blown skepticism is to be avoided, there might be some reason for moderate naturalists to be scientific realists after all.

Another way of putting the issue is to say that the "risk as little as possible" maxim is not only philosophically uninteresting in the same sense as skepticism,[8] but is also not consistently followed by antirealists, to the extent that they are less than genuine skeptics. On the other hand, the realist intuition is forceful, especially when it comes to explaining things like the unifying power of theories and, even more, their ability to *predict* new and unexpected phenomena – the latter, indeed, seem much more expectable in a realist context. Consequently, naturalists can, and perhaps ought to, follow the realist interpretation of science as an ever-improving product of natural evolution that provides us with more than mere instrumental capacities and practical results (it seems to me that Kitcher's claims that correspondence truth is rooted in everyday practice and that van Fraassen's observability criterion is refuted by science itself (Kitcher 1993, 130–131, 155), as well as Giere's claim that van Fraassen's views on science are themselves empirically inadequate (Giere 1988, ch. 5, § 5), go exactly in this direction).

Summing up this section: a restrictive conception of naturalism, in line with the tenets of logical empiricism, may significantly reduce the autonomy of philosophical inquiry, setting aside many issues, including that of scientific realism versus antirealism. However, a more moderate version of naturalism – one granting that (as in science) observation and experience do not do all the work, and philosophy must be continuous with science but need not be entirely reducible to it – might instead enable one to say more. In particular, moderate naturalism might enable one to preserve the philosophical significance of the scientific realism debate in spite of the fact that the opposing views are grounded in conflicting intuitions, rather than on firm empirical evidence. In addition to this, if inference to the best explanation and the role of truth and existential claims in philosophical explanations are granted, the realist intuition gains some non-negligible support – on the basis of considerations having to do with prediction, unification, explanatory power, and the continuity between everyday objects and experience, on the one hand, and the things

7 If I can only be certain of the contents of my sensorial perceptions, the reasoning goes, then it is already a risky theoretical leap to draw an inference from these to the existence of persisting objects and events that cause them.
8 If one wants to be a skeptic – setting aside the difficulty of being a *consistent* skeptic – then the problem arises that, while it is true that antiskeptic arguments satisfying one's epistemic standards are definitely hard to come by, there is little (if anything) more for one to do than simply state the inevitability of skepticism (given those epistemic standards).

and processes that are involved at the more sophisticated level of scientific inquiry, on the other. (For an analogous defense of the relevant forms of inference to the best explanation and of the idea that scientific realism is "scientific but not purely scientific," see Alai 2012).

Naturalism, Metaphysics, and Science

At least for the sake of argument, let us now assume scientific realism. The question that then emerges is what we take the (approximate) truth of scientific theories to amount to. As indicated earlier on, one might make the *epistemic* claim that science can give us true knowledge, while not making any *metaphysical* claim about exactly what sorts of thing it leads us to believe exist out there. For instance, one might believe that nonrelativistic quantum mechanics is approximately true, and closer to the truth than classical mechanics, without also taking a stance with respect to the important interpretational issues raised by the theory. (Are quantum particles individuals? How should we conceive of quantum properties? What does the measurement problem tell us about reality?) However, it must be noted that this position seems to be inherently unstable, especially in the case of certain "selective" forms of scientific realism. For instance, the previously mentioned entity realists should certainly say *what* entities exist out there (i.e., which existential claims of certain theories are true), and the same goes for those versions of realism that give a privileged status to, say, properties or relational structures. Thus, even conceding that the more generic forms of realism that focus on science as a whole, or on entire disciplines or theories, might avoid specifying their ontological commitments, this is not so in all cases.

If the foregoing is correct, however, then it is legitimate to ask on which basis one should limit oneself to the more general and generic form of scientific realism, and refrain from making any additional commitment. In particular, why should one be entitled to make a commitment to the existence of certain unobservable entities and processes as they appear in the "vocabulary" of scientists (e.g., subatomic particles), but not to others that are typically employed by philosophers, in particular metaphysicians, and which seem to play a fundamental role in the interpretation of scientific theories (e.g., objects, properties, individuals, etc.)? In fact, one might wonder whether a definition of the ontological commitments of a scientific theory (expressed via the first vocabulary) does not in fact depend on certain purely philosophical concepts (expressed in the other vocabulary). In a word, at least in a broadly scientific realist context, a question arises – and indeed, appears particularly pressing for the naturalist – concerning the limits of ontological commitment. And this immediately leads to questions concerning the status of metaphysics, both (1) in contrast with science and (2) as a potential basis for an interpretation of scientific theories.

In this case, too, a continuum of positions can be identified. As expected, the rejection of metaphysics as a basis for further analysis of scientific ideas, concepts, and findings is pretty strong within the empiricist tradition, and an eliminativist answer to these questions can easily be found there. Van Fraassen (2002), for instance, argued that while metaphysical questions are not meaningless, they are useless. Science, on the one hand, is constantly and harshly tested, and often falsified; this grounds its practical relevance. Metaphysics, on the other hand, seeks the truth, but is never in a position to establish whether what it says is actually true or false; thus, metaphysics is a merely formal exercise. Additionally, van Fraassen claimed that metaphysical questions are irredeemably context-dependent, and are such that they lack well-defined "answering strategies." The ensuing, final objection, crucial

for our present purposes, is that metaphysics accounts for "what we initially understand [in terms of] something hardly anyone understands" (2002, 3), and consequently turns out to be a superfluous addition to the indications coming from empirical science.

These arguments clearly converge toward a deflationary form of naturalism, whereby metaphysical inquiry loses relevance. The resulting perspective, of course, meshes perfectly with van Fraassen's constructive empiricist attitude. Since constructive empiricists already don't attach truth value to claims about what is not observable in the case of undisputedly scientific claims and hypotheses, they will *a fortiori* take all claims about, say, universals or object stages (metaphysical non-observables) as not worth thinking about. Analogously, the rejection of metaphysics is in line with the Fine–Ritchie agnostic take on the scientific realism issue.

Can a more moderate naturalist, one who is also a scientific realist, say something different? First of all, the idea of clearly demarcating science and metaphysics has proved to be a failure, it is normally accepted that there is a metaphysical element in virtually every piece of scientific reasoning and theorizing (most probably, the converse also holds, especially from a naturalistic viewpoint). But let us set this aside and focus, rather, on van Fraassen's objections against metaphysics. Taking my cue from his claim that metaphysics inevitably makes things less clear rather than clearer, I now wish to suggest that metaphysics is fundamental in interpreting scientific theories, making them more understandable in terms of general categories and concepts that build a bridge, so to speak, between what Sellars called the "manifest image" and the "scientific image" of the world.[9] For example, some views about the nature of objects and their properties have been developed with a view to solving some of the interpretative problems that arise when one deals with quantum mechanics (for more on the issue of individuality in quantum mechanics and science more generally, see Dorato and Morganti 2013). But these views and concepts belong to a nonscientific domain, the domain of metaphysics, normally inquired into via *a priori* means. The same holds for the nature of space and time in the context of relativity theories, for the nature of individuality and personal identity in biology, for the nature of chemical properties and natural kinds, and so on. It thus follows that metaphysical concepts can be, and in fact are, used to shed light on scientific theories and the descriptions of the world that the provide. Thus, the claim that metaphysics does not understanding, and in fact obscures things, appears completely unwarranted – unless a purely instrumental sense of understanding is presupposed, which would however be weak and question-begging. In addition to this, it is also false that metaphysics is merely a conceptual game with no clear rules and guiding methodology. First, one may speak of a form of "indirect testability," whereby metaphysical concepts and hypotheses are responsive to empirical input, at least to some extent, exactly in virtue of their being employed to interpret our best scientific conceptions of that input. That is, it is their ability to ground the interpretation of science that provides the naturalist with general "strategies" for evaluating metaphysical hypotheses. In connection to this, note that non-eliminativist naturalists about metaphysics need not show that

9 See, for instance, Sellars (1962). According to Sellars, the commonsense view of reality (according to which, for instance, tables and other middle-sized objects undoubtedly exist) and the more sophisticated scientific picture of things (according to which, for instance, macroscopic things are nothing but complex structures of elementary particles, fields, or what have you) are normally in tension with one another. However, says Sellars, a "stereoscopic view," somehow establishing a peaceful coexistence between the two, can and should be sought. We will not discuss the plausibility of this irenic view here.

"indirect testability" provides us with conclusive and unambiguous indications as to the relative strengths of the different available metaphysical hypotheses. Indeed, no such indications should be expected, for there certainly is no mechanical way of determining the "degree of fit" of these hypotheses with the empirical data as codified by our best theories; and it is likely that non-epistemic factors such as the conservation of established beliefs, the connection with other hypotheses, and simplicity play a role. But exactly these same difficulties and pragmatic aspects of theory-choice are normal in science! Hence, considerations of continuity between science and nonscience, perfectly fitting the naturalist's assumptions and general methodology, appear to tip the balance against the more radical eliminativist/deflationist positions.

There is, however, a third option that must be taken into account. Ladyman and Ross (2007), for example, condemn contemporary *a priori* metaphysical inquiry (which they call "neo-Scholastic"), urging metaphysicians to engage with science directly. Against this background, however, rather than opting for eliminativism with respect to metaphysics, they put forward a positive metaphysics (one based on structures and reality being, at its basis, a network of relations), which they seem to regard as "extractable" from our best science – *in primis* physics – more or less directly. How compelling is their account?

First of all, a key idea underlying Ladyman and Ross's argument seems to be that the autonomy that one might grant metaphysics depends on there being something like a realm of metaphysical possibility and necessity that outstrips the possibility and necessity that science deals with, but this is exactly what naturalists should not be willing to concede. If the only indication of what is possible or necessary comes to us from science, however, it would appear indeed correct to think that we should bypass metaphysics even for the specific purposes of laying the basis for the interpretation of scientific theories. Hence, Ladyman and Ross conclude, the vacuity of the notion of metaphysical possibility makes naturalism incompatible with the idea that metaphysics has some (albeit limited) degree of independence with respect to science – which in turn leads one to question moderate naturalism itself.

While a detailed discussion of this exceeds the scope of this paper (for more details, see Morganti 2013, where the role played by the notion of "metaphysical grounding" in clarifying issues concerning the interpretation of science is also emphasized), there are rejoinders available. To begin with, it is true that philosophers have often been wrong in claiming that something is possibly/necessarily (not) the case, and that the contrary has turned out to hold as science has progressed – Ladyman and Ross use as an example Kant's view of Euclidean geometry as *a priori* true. But hasn't the same happened in science, as witnessed by the many examples that antirealists are quick to provide? Think, for instance, of the mechanistic idea that light had to be propagated by an elastic medium. Fallibility surely cannot be a factor in a naturalistic context. More importantly, it may well be the case that when we recognize (or think we recognize) something as possible or impossible, we do so in a way that is affected by our experience and knowledge of the actual world. But at no point is the (naturalized) metaphysician forced to deny this: the weaker claim is sufficient that there is *something* that is provided by *a priori* metaphysical analysis that cannot be found "already there" in scientific theories. For this to obtain, it is sufficient that questions can be meaningfully asked that science leaves (necessarily or merely contingently: it is conceivable, if not plausible, that at least some metaphysical questions turn into scientific ones as research progresses) underdetermined; and this in turn requires only that there be peculiar categories and concepts of metaphysics, different from those of science and not

definable in terms of the latter – *not* that there be a specific, *sui generis* domain of possibility/necessity that can only be inquired into by metaphysics. Indeed, one may plausibly suggest that it is the use of precisely these concepts and categories, supposed to be applicable independent of the specific part or domain of reality that one is observing, that grounds the status of metaphysics as an autonomous discipline – irrespective of whether or not this coincides with there being a *sui generis*, irreducibly metaphysical, modal structure of reality.[10] (On the other hand, positive arguments can be, and have been, given in favor of pluralism about modality – see, for instance, Fine (2002). Therefore, a neutral stance with respect to the reducibility or irreducibility of metaphysical to physical/nomological modality might be regarded as the best choice for moderate naturalists.)

Another thing must be pointed out before closing this section. As mentioned earlier, Ladyman and Ross launch an attack against traditional metaphysics but also put forward a positive metaphysics (of structures) of their own. How is this proposal to be understood? It seems implausible that a metaphysics of structures can be deductively inferred from our best scientific theories. And it also seems that the concepts used in formulating such a metaphysics (object, relation, identity, property, etc.) do not belong to the vocabulary of, say, high-energy physics. Doesn't this mean that the proposal being considered relies on a prior exploration of exactly the *sui generis* space of metaphysical possibility that Ladyman and Ross are eager to put into question?

To wrap up, a moderate form of naturalism is possible, and indeed advisable, in the case of the science/metaphysics relationship, too. Naturalism about metaphysics, in particular, is usefully portrayed as the view that metaphysics identifies (by making use of *a priori* methods of logical and conceptual analysis, and employing *sui generis* linguistic/conceptual categories) the possibilities that are relevant for the interpretation of scientific theories; and, in turn, the use of these peculiar tools for interpreting science makes it possible to further evaluate the metaphysical possibilities themselves – thus starting a potentially virtuous process of mutual support and parallel development.

Conclusion

Before closing, I would like to repeat some key points and suggest at least in broad terms the features of a unified, overarching moderate naturalistic picture in the philosophy of science broadly understood.

First, naturalistic methodology can play, and has played, a role in the philosophy of science, primarily with respect to issues concerning the epistemic import of scientific theories. This role is grounded in the conception of scientific theories as a means for representing and interacting with the world, in continuity with our natural cognitive apparatuses. Among

10 Indeed, the traditional definition of metaphysics – according to which metaphysics deals with being in its utmost generality – seems best understood in the sense that one employs the most general concepts and categories, and not in the (quite widespread) sense that one deals with the widest possible domain of inquiry. After all, the "Theory of Everything" dreamt of by many physicists doesn't count as metaphysical just because it is the most general theory one might hope to formulate. Metaphysics, instead, uses – and has consistently used throughout its history – concepts and categories that are different from, irreducible to, and more fundamental than those of the special sciences. And such irreducibility seems to be due exactly to the fact that these are the most general concepts and categories (for instance, those of object, property, and individuality mentioned earlier, or the notions of substance, causation, etc.), regardless of how "locally" they are applied.

other things, it entails that we ought to look at science as a real-world dynamic entity shaped by epistemic as well as non-epistemic goals and factors.

Second, strong naturalism of the empiricist and neopositivist varieties leads to radical, agnostic/deflationist stances with respect to both scientific realism and the possibility of metaphysics. However, a more moderate naturalism is also possible, whereby (1) to naturalize does not mean only to accept what lends itself to direct empirical verification, but does require *some* form of continuity between science and metaphysics, and (2) the acceptance of inference to the best explanation and of beliefs that are less than 100% certain (which seem to be common currency in science) legitimizes a broader range of deeply interesting philosophical issues. Such a perspective should appear clearly preferable to all those who feel the pull of the naturalist intuition and yet would like the scope of the philosophical questions worth asking to be limited – and, hence, philosophy itself to be diminished – as little as possible.

Third, I suggested that moderate naturalists might have reasons for subscribing, on the one hand, to a form of scientific realism (though one that is likely to require a case-by-case evaluation on the basis of the available data) and, on the other, to a view of metaphysics as an enterprise worth pursuing in spite of its essentially *a priori* nature – provided that it be made to interact systematically with our best science. In particular, I suggested that this interaction should be based on the identification of possible ways of conceiving of certain parts of the world, the use of these conceptions for the interpretation of scientific theories, and the critical assessment of the metaphysical hypotheses themselves in light of this application.

Much more needs to be said about these matters, of course. But in any case, moderate naturalism – in the form sketched here – definitely appears to be a coherent position in the philosophy of science, and one worth giving further consideration. In particular, it seems to have the unique advantage of satisfying the sensible demand for awareness of the status of science on the part of philosophers, while at the same time avoiding the deflation of philosophy in terms of questions worth asking and conceptual tools that can be used to provide answers – something that has often occurred in the past, on the basis of very general (philosophical!) assumptions that were both questionable and put into practice in a non fully-consistent manner.

Whatever one thinks of the more exploratory claims that have been made concerning moderate naturalism and its prospects, I hope at any rate to have provided here a useful general overview of topics and issues related to naturalism and the philosophy of science.

References

Alai, M. (2012). Levin and Ghins on the "No Miracle" Argument and Naturalism. *European Journal for the Philosophy of Science* 2: 85–110.

Bloor, D. (1976). *Knowledge and Social Imagery*. London: Routledge and Kegan Paul.

Boyd, R. (1996). Realism, Approximate Truth and Philosophical Method. In *Philosophy of Science*, edited by David Papineau, pp. 215–255. Oxford: Oxford University Press.

Dorato, M. and Morganti, M. (2013). Grades of Individuality. A Pluralistic View of Identity in Quantum Mechanics and in the Sciences. *Philosophical Studies* 163: 591–610.

Fine, A. (1984). The Natural Ontological Attitude. In *Scientific Realism*, edited by Jarrett Leplin, pp. 83–107. Berkeley: University of California Press.

Fine, K. (2002). The Varieties of Necessity. In *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne, pp. 253–282. Oxford: Oxford University Press.

Giere, R. (1988). *Explaining Science. A Cognitive Approach*. Chicago and London: University of Chicago Press.

Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.

Hempel, C.G. (1966). *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall.

Howson, C. (2000). *Hume's Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.

Kitcher, P. (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Ladyman, J. and Ross, D. (with Spurrett, D. and Collier, J.) (2007). *Every Thing Must Go. Metaphysics Naturalised*. Oxford: Oxford University Press.

Lakatos, I. (1971). History of Science and its Rational Reconstructions. In *Boston Studies in the Philosophy of Science*, edited by Roger C. Buck and Robert S. Cohen, pp. 91–135. Dordrecht: Reidel.

Latour, B. and Woolgar, S. (1979). *Laboratory Life*. Beverly Hills: Sage.

Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science* 48: 19–49.

Laudan, L. (1987). Progress or Rationality: The Prospects for Normative Naturalism. *American Philosophical Quarterly* 24: 19–31.

Maddy, P. (2007). *Second Philosophy*. Oxford: Oxford University Press.

Magnus, P.D. and Callender, C. (2004). Realist Ennui and the Base Rate Fallacy. *Philosophy of Science* 71: 320–338.

Morganti, M. (2013). *Combining Science and Metaphysics: Contemporary Physics, Conceptual Revision and Common Sense*. Basingstoke: Palgrave Macmillan.

Price, H. (2004). Naturalism without Representationalism. In *Naturalism in Question*, edited by Mario De Caro and David Macarthur, pp. 71–88. Cambridge, MA: Harvard University Press.

Putnam, H. (1978). *Meaning and the Moral Sciences*. London: Routledge.

Ritchie, J. (2008). *Understanding Naturalism*. Durham: Acumen.

Sellars, W. (1962). Philosophy and the Scientific Image of Man. In *Frontiers of Science and Philosophy*, edited by Robert Colodny, pp. 35–78. Pittsburgh: University of Pittsburgh Press.

van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

van Fraassen, B. (2002). *The Empirical Stance*. New Haven: Yale University Press.

7

Naturalism without Scientism

P. KYLE STANFORD

Integrative Naturalism

Let us begin with naturalism itself. These days, most philosophers seem eager to regard themselves as naturalists, but this is possible only because there is so little agreement about what naturalism actually involves or requires. There is, of course, a straightforwardly (and self-consciously) scientistic form of philosophical naturalism, which simply assumes that the claims of our best scientific theories should serve as both starting point and boundary conditions for any serious form of inquiry. Alexander Rosenberg, for instance, boldly announces that "Naturalism is the label for the thesis that the tools we should use in answering philosophical problems are the methods and findings of the mature sciences – from physics across to biology and increasingly neuroscience" (2014, 17). Likewise, according to the "radically naturalistic metaphysics" defended by James Ladyman and Don Ross (2007, 1), "Naturalism requires that, since scientific institutions are the instruments by which we investigate objective reality, their outputs should motivate all claims about this reality, including metaphysical ones" (2007, 30). Like Rosenberg, Ladyman and Ross cheerfully plead guilty to the charge of "scientism" (2007, 65), and they insist that metaphysical claims should not even be taken seriously unless they seek to unify the independent findings of our best scientific theories:

> Any metaphysical claim that is to be taken seriously should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses jointly explain more than the sum of what is explained by the two hypotheses taken separately, where a "scientific hypothesis" is understood as an hypothesis that is taken seriously by institutionally *bona fide* current science. (2007, 30)

Philosophers and other nonscientific inquirers are thus advised to reform their investigations so as to begin from the results and methods of institutionally *bona fide* scientific inquiry itself, or at the very least not stand in the way of making further progress in such

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

inquiry; as the song says, "get out of the way, if you can't lend your hand, for the times they are a-changin'."

Though combining such ardently scientistic varieties of philosophical naturalism with resistance to scientific realism might indeed be incoherent or a contradiction in terms, I see no compelling reason to embrace their dogmatic starting point. It is, of course, true that any satisfying form of naturalism will have to make good sense of the dramatic and undeniable practical achievements of our best scientific theories, but a more appealing and systematic motivation for this demand can be found in W.V.O. Quine's (1951/1980) suggestion that there is only a single, interconnected project of inquiry into the world and our own place within it, and his insistence that this single integrated inquiry is responsible for simultaneously making good sense out of *all* of the reliable evidence we have available to us at any given time. Such evidence will certainly include the impressive feats of prediction, control, and putative explanation achieved by our best scientific theories, but it will include much more in addition. This *integrative* form of naturalism will require that we also make sense of the similarly impressive practical and explanatory accomplishments of many past successful scientific theories that we now regard as offering quite profoundly mistaken accounts of the fundamental constitution of some part of the natural world, of the radical discontinuities between succeeding generations of empirically successful theories in the same scientific domain across the history of science, of our repeated failure to even conceive of scientifically serious alternative theoretical possibilities that were well confirmed by the evidence available at the time they remained unconceived, and of much else besides. A convincing integrated conception of ourselves and our place in the world must be no less responsive to such historical evidence than it is to the impressive experimental demonstrations, explanatory achievements, and other sorts of empirical evidence we have found to support contemporary scientific theories themselves.

As Quine suggests, the naturalistic philosopher pursuing such an inquiry

> begins his reasoning within the inherited world theory as a going concern. He tentatively believes all of it, but believes also that some unidentified portions are wrong. He tries to improve, clarify, and understand the system from within. He is the busy sailor adrift on Neurath's boat. (1975/1981, 72)

Such a Quinean naturalist will not *begin* her inquiry by privileging any part of her "inherited world theory" – her efforts to "improve, clarify, and understand the system from within" are responsible for making sense of the entire body of empirical evidence she takes herself to have at any given time. The only truly consequential distinction between *kinds* of evidence that such a naturalist will recognize is between *good* and *bad* evidence, and even that division will have to be earned from within the ongoing project of inquiry itself.

At the heart of any appealing naturalistic conception of inquiry, I am suggesting, is neither a dogmatic scientism nor even Quine's own famously overzealous holism about testing or meaning, but instead simply the broadly Quinean insistence that *all* (good) evidence matters, and a consequent refusal to automatically elevate scientific, historical, philosophical, or any other particular form of evidence into a privileged position while simply neglecting the rest. On this view, evidence concerning the distinctive characteristics of our scientific theories and ourselves as theorizers must feed back to inform our view of what we are doing when we theorize about the world, of what we can and cannot reasonably expect such scientific theorizing to accomplish under various sorts of epistemic conditions,

and of the character, reach, and epistemic status of the scientific theories generated by this process. For such integrative naturalists, understanding what our best scientific theories are telling us about the world and understanding how we go about entheorizing that world in the first place are not distinct challenges: both are part of the overarching and more fundamental challenge of trying to *simultaneously* understand both the world and our own place within it.

I expect many philosophers of science to find this conception of naturalism congenial. After all, those of us who think of what we do as broadly continuous with the sciences often do so *because* we think that there is only a single project of inquiry and that we are engaged in it jointly with scientists themselves. That project has plenty of moving parts, from developing new statistical methods and quantitative measures of confirmation to building better microscopes, and from proposing and testing novel empirical hypotheses to interpreting the historical record of scientific inquiry, but these are all parts of a single interconnected inquiry into both the world and our own place in it, including the place of activities like science itself within that single project of systematic inquiry.[1]

Indeed, this conception of naturalism might even seem too ecumenical, insofar as many paradigmatically non-naturalistic philosophers will presumably be happy to sign on to *some* version of the idea that all of our efforts to understand the world and our own place in it are simply parts or aspects of a single, integrated inquiry. It is therefore worth noting explicitly that this integrative naturalism will reject the persistent suggestion that there are distinctively philosophical modes of inquiry or characteristically philosophical routes to knowledge that are categorically distinct from those offered by our ordinary practices of empirical and scientific investigation. The suggestion that there are such distinctively philosophical forms of inquiry (sometimes alleged to have a stronger claim on our credence than ordinary empirical methods of investigation) has played a prominent role across the history of the discipline, whether we think of Plato's prescriptions for retrieving our souls' memories of the Forms carried from before our births, Descartes' clear and distinct perception of his own ideas or concepts, or Kant's transcendental inquiry into the conditions of possible experience, to name just a few historically influential examples.

To be sure, those who defend such distinctively philosophical forms of inquiry have usually offered some rationale for thinking that they can deliver important and otherwise inaccessible truths concerning nature and/or our own place within it. Descartes, for example, thought he had decisive reasons for regarding his concepts as the creations of an omnipotent, omniscient, and benevolent God who would (therefore?) not deceive him so long as he was careful to reason from and about those concepts judiciously. And Kant's "Copernican revolution" is widely seen as insisting that much of the structure of the external world as we experience it is imposed by the fundamental concepts we use to transform the raw materials available to us in sensation into genuine experience of such a world in the first place, and thus that the investigation of those concepts reveals how the world must be for us.[2]

1 This proposal closely parallels the much more detailed form of philosophical naturalism influentially defended by my colleague Penelope Maddy (see esp. 2007), and I am deeply indebted to her for many insights and useful conversations on this subject. Despite our agreement concerning the evidence to which an appealing form of naturalism must be responsive, however, she would sharply dissent from many of the conclusions I will go on to draw from that evidence.

2 There are, of course, other profound differences between Descartes and Kant, but the relevant similarity is that each is at least sometimes willing to set all empirical investigation aside and try to reach substantive conclusions about ourselves and our place in the world on the basis of some categorically distinct form of inquiry.

The integrative naturalist I have described will, of course, refuse to simply *presume* that there are any such independent routes to knowledge in advance of inquiry, and the presently available evidence will lead her to regard existing rationales for such distinctively philosophical forms of inquiry as manifestly unpromising. It is worth recalling that Kant's own case for the permanence of the fundamental Categories, for example, rested in large part on the fact that the logical apparatus from which they were generated had not changed substantially since the time of Aristotle – but Kant made this argument only a hundred years before Frege radically transformed the logic whose presumed permanence was the ground on which Kant thought he could identify such metaphysical fixed points in the ways we are capable of experiencing and/or cognizing the world. Though there may well be obligate features of the ways human beings are able to experience or conceptualize the world, empirical psychology now seems a far better bet for identifying whatever we can know about them than any distinctively philosophical form of transcendental inquiry. Of course, contemporary philosophers sometimes suggest that conceptual and/or linguistic analysis can ground such independent inquiry because it is only by satisfying our concepts and/or the meanings we assign to the associated terms that something can ultimately *count* as a "material object" or an "injustice." But such conceptual and/or linguistic analysis seems considerably better suited to probing how we (presently) think and talk about the world than to providing substantive information about how the world itself is or must be (even how it is or must be *for us*). Though the specific intuitions, concepts, and linguistic judgments we find ourselves with at any particular time do, of course, themselves constitute a sort of empirical data, the balance of the evidence we have concerning the sources and plasticity of such intuitions, concepts, and judgments offers little reason to suppose that they offer distinctively philosophical routes to substantive knowledge of ourselves and the world around us that are independent of our broader empirical and scientific inquiries.

This integrative conception of philosophical naturalism clearly owes an enormous debt to Quine: at its heart are such central Quinean insights as the fact that there is only a single, integrated project of inquiry into the world and our place within it, that this inquiry must seek to improve the "inherited world theory as a going concern" from within without automatically privileging some reliable sources of evidence over others, and that it will neither presuppose nor conclude at present that there are any distinctively philosophical routes to substantive knowledge about ourselves or the world. If all this is right, however, Quine himself stumbles badly when he famously suggests that epistemology should simply be assimilated to or replaced with empirical psychology:

> Naturalism does not repudiate epistemology, but assimilates it to empirical psychology. Science itself tells us that our information about the world is limited to irritations of our surfaces, and then the epistemological question is in turn a question within science: the question how we human animals can have managed to arrive at science from such limited information. (1975/1981, 72)

This Quinean impulse has been criticized in a wide variety of ways, but the most fundamental mistake here *concerning naturalism* would seem to be the suggestion that evidence concerning how sensory irritations generate scientific beliefs will be all the evidence that ultimately matters to epistemology. The central concern of epistemology is when and how we are *justified, warranted, or well-advised* to hold particular beliefs or sorts of beliefs in particular sorts of epistemic circumstances (and why), and the thoroughgoing integrative

naturalism I have described will see much more than just the evidence supporting the claims of our best psychological theories concerning the details of the causal chains between the "irritations of our surfaces" and scientific beliefs themselves as relevant to making that determination.[3] More specifically, it will insist that such judgments must be responsive to evidence from the history of science, existing scientific practices, and much else that goes beyond the traditional purview of empirical psychology.

Epistemic Instrumentalism

Suppose now that we do replace the scientistic naturalist's exclusive focus on the impressive empirical achievements of our best scientific theories with a much wider conception of the evidence to which a satisfying integrated conception of the world and our own place within it must be responsive. Why will this make any difference to the view we take of the descriptions offered by those theories of how things stand in otherwise inaccessible domains of nature or challenge the scientific realist's conviction that those descriptions must be at least probably and approximately true?

Critics of scientific realism have long noted that in most scientific fields contemporary theories are simply the latest in a succession of fundamentally distinct theoretical alternatives, many of which enjoyed impressive empirical successes of at least the same broad kinds as those enjoyed by their contemporary successors. Some critics (e.g., Laudan 1981) argue that this long history of empirically successful scientific theories now regarded as profoundly mistaken in their fundamental claims about the constitution of nature (including such famous examples as Newton's mechanics, the phlogiston theory of chemistry, the caloric theory of heat, the wave theory of light and electromagnetism, Weismann's theory of the germ-plasm, and many more besides) establishes that even the most dramatic forms of empirical success are simply not a reliable indicator that a scientific theory is probably and/or approximately true. Others (e.g., Stanford 2006) suggest that our repeated failures to even conceive of the fundamentally distinct and even more empirically impressive successors that would ultimately come to replace such theories shows that as inquirers we are not proficient at conceiving of the full range of theoretical alternatives that are well confirmed by a given body of empirical data, and thus that there are very likely alternatives to contemporary scientific theories that remain unconceived by us despite the fact that they are also reasonably well confirmed by the evidence we now possess. It is to this sort of historical evidence concerning the reach of our inquiry and our own capacities as inquirers that the integrative sort of naturalist I have described will insist that our conception of the world and our own place within it must be no less responsive than it is to the dramatic empirical achievements that convince scientific realists that successful contemporary scientific theories must be at least probably and/or approximately true.

I suggest, then, that even if such a naturalist begins her inquiry from within our "inherited world theory as a going concern," she will ultimately come to see the *full range* of evidence

3 Quine himself denies that naturalizing epistemology in this way sacrifices its normative dimension (1981, 181; 1990, 19; 1995, 49; see also Maddy 2007, § I.6 for useful discussion). But what matters most here is why the integrative naturalist will resist the sort of assimilation Quine *appears* to endorse in this and other passages: the fundamental concern is that Quine's own naturalism takes too narrow a view of the range of relevant evidence to which our epistemological judgments should be sensitive. I will go on to say more in later sections about how expanding the range of such evidence contributes to recovering the normative dimension of epistemology.

informing our integrated conception of the world and our own place within it as indicating that we lack sufficient warrant for regarding some of even the most successful contemporary scientific theories as probably and/or approximately true. She will resist the scientific realist's conception of such theories, despite the fact that she cannot propose any alternatives to them that she thinks have better or even equally good claims to represent the truth concerning the relevant otherwise-inaccessible domains of nature, and even if she cannot now say anything about what she ultimately expects their still more empirically impressive and fundamentally distinct successors to look like. But having reached that conclusion, what view will our integrative naturalist take of such theories instead? Where and how will they fit into the inherited world theory that she seeks to improve from within?

The most historically influential alternative to the realist's view that our best scientific theories are at least probably and/or approximately true has been the competing proposal that such theories are instead "mere instruments," which is to say useful cognitive and conceptual tools that we use to navigate our practical predicaments and to predict and intervene in the world around us as successfully as possible. On such a view, a scientific theory is no less a tool for being conceptual in character, or a "way of thinking" about some otherwise inaccessible domain of nature. Hilary Putnam famously pointed out (in a different context) that there are tools like a hammer, which require only one person to use, and tools like a steamship, whose effective use requires the coordinated efforts of many people. Here we might simply add that many of the tools we use, like mnemonic devices, work schedules, and recipes, are not physical objects at all.

In the middle decades of the 20th century, many defenders of instrumentalism famously embraced *semantic* versions of this proposal, variously holding that the apparently theoretical claims of science are really just claims about observable phenomena in disguise, that such claims do not actually assert anything at all but instead simply licensed inferences from one observable state of affairs to another, or that such claims are in principle eliminable from scientific discourse altogether. For good reason, each of these semantic proposals is now widely regarded as facing insurmountable obstacles (cf. Stanford 2005). But the fundamental idea that at least some of even our best scientific theories should be viewed merely as conceptual tools or instruments for guiding our practical engagement with the world in productive and instructive ways, rather than as accurate descriptions of how things stand in otherwise inaccessible domains of nature, depends *in no way* on the demand for a special semantic interpretation or translation of scientific claims. We might instead embrace an *epistemic* version of such instrumentalism, on which we see scientific theories as *saying* just what they seem to say about what the world is made up of and how it works, but nonetheless deny that we have sufficient justification for *believing* those claims when they are so understood. Such epistemic instrumentalism holds that scientific theories frequently can and do help us successfully navigate the world in productive and systematic ways, even when the claims they make about otherwise inaccessible domains of nature are not even approximately true. Accordingly, the epistemic instrumentalist thinks that we can and should *use* those theories to guide our pragmatic engagement with the world without *believing* what they say about how things stand in otherwise inaccessible domains of nature.

Such epistemic versions of instrumentalism face a persistent and powerful line of criticism, however, to the effect that there is no room for a genuine difference between making "merely" instrumental use of a scientific theory across the *full* range of diverse contexts in which such theories are put to work and simply believing such a theory to be true. Howard Stein argues, for example, that once we recognize that the instrumental uses of a theory

include not only calculating experimental outcomes but also adequately representing phenomena more generally and serving as "resources for inquiry; especially as sources of clues in…the search for good hypotheses" (1989, 52), there simply remains no room for a difference that *makes* any difference between such a thoroughgoing instrumentalism and a suitably modest form of scientific realism that gives up any claim to be telling us about how things stand independent of the ways it is possible for us to experience and/or conceive of them.

Simon Blackburn offers a similarly motivated but more detailed criticism of Bas van Fraassen's influential "constructive empiricist" form of epistemic instrumentalism, which famously holds that instead of believing even our best scientific theories, we should merely "accept" them as "empirically adequate," meaning we should believe all and only what they say about observable phenomena and remain agnostic about their further claims regarding the unobservable. But as Blackburn notes, the constructive empiricist nonetheless recommends that we "immerse" ourselves in our theories in order to take full advantage of them:

> The constructive empiricist is of course entirely in favor of scientific theorizing. It is the essential method of reducing phenomena to order, producing fertile models, and doing all the things that science does. So we are counseled to *immerse* ourselves in successful theory…Immersion will include acceptance as empirically adequate, but it includes other things as well. In particular it includes having one's dispositions and strategies of exploration, one's space of what it is easy to foresee and what difficult, all shaped by the concepts of the theory. It is learning to speak the theory as a native language, and using it to structure one's perceptions and expectations. It is the possession of habits of entry into the theoretical vocabulary, of manipulations of its sentences in making inferences, and of exiting to empirical prediction and control. Van Fraassen is quite explicit that all of this is absolutely legitimate, and indeed that the enormous empirical adequacy of science is an excellent argument for learning its language like a native… Immersion, then, is belief in empirical adequacy plus what we can call being "functionally organized" in terms of a theory. (2002, 117–119; emphasis in original)

But once we are "immersed in" or "animated by" our best scientific theories in order to take full instrumental advantage of them, Blackburn suggests, we have lost any room for a distinction between regarding our theories merely as useful cognitive instruments and simply believing them to be true:

> The problem is that there is simply no difference between, for example, on the one hand being animated by the kinetic theory of gases, confidently expecting events to fall out in the light of its predictions, using it as a point of reference in predicting and controlling the future, and on the other hand believing that gases are composed of moving molecules. There is no difference between being animated by a theory according to which there once existed living trilobites and believing that there once existed living trilobites…What can we do but disdain the fake modesty: "I don't really believe in trilobites; it is just that I structure all my thoughts about the fossil record by accepting that they existed"? (2002, 127–128)

Like Stein, Blackburn suggests that recognizing the full range of instrumental uses of a theory seems to eliminate any room we might have to distinguish "merely" instrumental reliance from belief itself. Thus, as we progressively sophisticate our grasp of the various ways in which we make use of our best scientific theories or put those theories to work, it becomes harder and harder to see how such systematic reliance on or use of a scientific

theory amounts to anything less than simply believing that same theory. The epistemic instrumentalism embraced by our integrative philosophical naturalist thus threatens to simply collapse into scientific realism itself!

We can see that something has gone badly wrong in this line of argument by noting that the idea of making systematic instrumental use of a scientific theory we do not believe to be even approximately true turns out to be one that the scientific realist needs no less than the epistemic instrumentalist – after all, this is precisely the view that the realist herself takes of a theory like Newtonian mechanics! That is, the scientific realist believes the claims of Newton's mechanics about the fundamental constitution and operation of nature to be profoundly mistaken: gravitational motion is not caused by forces exerted by massive bodies on one another, space and time are not absolute, and so on. But she knows perfectly well how to use Newton's mechanics rather than Einstein's to send rockets to the moon and in a wide array of other engineering applications in which we know that any differences in their empirical predictions are small enough to be safely disregarded. Thus, we might begin by asking what the *realist* means when *she* says that we make effective instrumental use of Newtonian mechanics to solve practical problems and challenges but we do not believe its fundamental claims about the constitution of nature to be even probably or approximately true.

The most natural way to describe the realist's attitude toward Newtonian mechanics would seem to be that though she does not believe the theory's claims about the fundamental constitution and operation of nature to be true, she nonetheless knows how to apply the theory as a true believer would to entities, events, and natural phenomena she thinks she can correctly characterize in a variety of other ways. That is, the realist accepts the existence of entities like rockets, rattlesnakes, and RNA molecules on the evidence of her senses and/or the descriptions offered of them by theories that she does strictly and literally believe, and she knows both how to represent such entities within a Newtonian framework and how to derive consequences and empirical predictions concerning their behavior when they are so represented. She knows how a Newtonian *would* characterize (in terms of forces, masses, etc.) the rockets, rattlesnakes, and RNA molecules familiar from her own experience and/or other theories that she takes to be probably and/or approximately true, and she can use those characterizations together with the dynamical apparatus of the theory itself to derive predictions, explanations, and recipes for intervention just as a Newtonian would concerning entities, events, and phenomena as she understands or conceives of them independently of Newtonian mechanics (or indeed any other theory toward which she herself adopts a similarly instrumentalist attitude).

Perhaps surprisingly, however, *this* conception of instrumentalism is equally available to the epistemological instrumentalist herself, so long as she recognizes at least some routes of epistemic access to entities, events, and phenomena in the world toward which she does not extend an instrumentalist attitude. Many classical instrumentalists have thought that objects and events *as they are given to us in sense perception* are those with respect to which scientific theories allow us to predict and intervene successfully. But the instrumentalist is by no means limited to the evidence of her senses in conceiving of the natural phenomena with respect to which she thinks a given scientific theory offers an effective instrumental tool for prediction, intervention, further investigation, or anything else: she is in a position to take full advantage of *whatever* routes of epistemic access she has to entities, events, and natural phenomena besides the descriptions of them offered in any scientific theories toward which she takes an instrumentalist attitude, no matter what those routes turn out to be. Indeed, this conception is available to the epistemic instrumentalist even if she embraces

Quine's further suggestion that *all* of our knowledge is theoretical in character and thus that even our understanding of the most familiar entities and events is given to us by a hypothesis of "the bodies of common sense" (1976, 250). So long as she does not adopt an instrumentalist attitude toward *this particular hypothesis*, the descriptions of entities and events it offers can serve perfectly well as the conception of those entities and events regarding which theories that are "mere instruments" will allow us to predict and intervene effectively. The instrumentalist, just like the realist, will simply be faced with a decision concerning *which* theories are those toward which she will adopt a realist attitude, and why all and only those.

This realization suggests that it was a mistake all along to think of the difference between the scientific realist and her instrumentalist opponent as a difference in the respective attitudes each takes toward "theories" or "theoretical knowledge" as such. Both realists and instrumentalists will take instrumentalist attitudes toward some theories (e.g., Newtonian mechanics) and realist attitudes toward others (e.g., the Quinean hypothesis of the bodies of common sense), and the crucial remaining difference between them is simply that thinkers who embrace the instrumentalist label are those who are prepared to treat a much *wider range* of theories merely as useful conceptual instruments for guiding our pragmatic engagement with and further investigation of nature than are their realist counterparts. In particular, the instrumentalist, but not the realist, will be prepared to adopt this attitude even toward theories that are clearly more empirically successful than any competitor we have yet managed to develop, even when she cannot yet say anything at all about what she expects their still more empirically successful and pragmatically powerful successors to look like.

Of course, scientific theories sometimes posit the existence of entities or events to which we have no route of epistemic access at all beyond the causal, dynamical, or other role ascribed to them by that very theory. Because contemporary particle physics does not allow quarks to be isolated, for example, it posits the existence of "gluons" to bind quarks within a proton, while contemporary cosmology posits the existence of dark energy to account for the accelerating expansion of the universe. In both cases, however, the ground for believing in these entities is simply the demand of the relevant theory that *something* play the role that each is hypothesized to fulfill. Someone taking an instrumentalist attitude toward contemporary particle physics (or cosmology) will therefore not believe in even the actual existence of gluons (or dark energy), much less any part of the theoretical description of them offered by the theory, but will *simply* make use of such descriptions and claims in seeking to successfully predict, intervene, further investigate, and so on with respect to *other* entities, events, and phenomena to which she does see herself as having independent epistemic access. That is, she will regard gluons or dark energy *simply* as instrumentally useful fictions, just as the realist so regards Newtonian gravitational forces or absolute space and time.

As this comparison suggests, plenty of room remains for reasonable disagreement concerning how far realist or instrumentalist attitudes should be extended to particular scientific theories. Though quantum mechanics is widely regarded as the most empirically successful theory in the history of science, for example, even many realists hesitate to embrace the idea that its descriptive claims about otherwise inaccessible domains of nature are at least probably and/or approximately true, in part because it seems so difficult to understand those claims *as* a description of what actually occurs in the physical world. In addition, quantum mechanics is straightforwardly inconsistent with other wildly successful physical theories in ways that it is not easy to see how we can avoid or overcome without making fundamental changes to one or more of those theories (cf. Barrett 2003).

More generally, there is ongoing disagreement concerning the right morals to draw from the historical record of scientific inquiry itself. Much controversy has centered on whether there are particular *characteristics* of some theories (such as their "maturity" or their ability to make successful predictions of novel phenomena) that should lead us to expect them to avoid the fate of their historical predecessors, and/or whether the historical record reveals particular *parts or aspects* of (sufficiently) successful scientific theories (such as their claims about the "structure" of nature, the entities they posit, the "working" posits to which they ascribe substantive causal roles, or those aspects regarded as best-confirmed by the relevant scientific communities themselves) to be invariably or at least reliably preserved by their historical successors, justifying our confidence that such theories, or at least their relevant parts or aspects, will persist throughout the course of all further inquiry. I argue elsewhere (Stanford 2006, chs. 6 and 7) that these rearguard maneuvers are unconvincing: the criteria used to pick out the relevant characteristics, parts, or aspects of theories either run afoul of notable counterexamples from the historical record or are framed in sufficiently vague and indeterminate ways as to provide little *prospective* guidance in picking out the relevant theories and/or parts or aspects of theories in which we can now safely invest our credence.

We need not resolve such disputes, however, to achieve the aim of the present chapter, which is simply to show how resistance to scientific realism *could* be combined with philosophical naturalism in an attractive and compelling way. We have already seen how widening the range of evidence to which our naturalistic inquiry must be responsive so as to include empirical evidence from the historical record and elsewhere might motivate the combination of an integrative philosophical naturalism with a thoroughgoing epistemic instrumentalism. Moreover, even if we ultimately decide that the real lesson of such historical evidence is simply that we should limit our belief to successful scientific theories that share some further characteristic (e.g., "maturity," the ability to make novel predictions), or simply to some parts or aspects of such theories (e.g., their "structure" or their "working" posits), this would already suffice to illustrate how the sort of integrative naturalism I have proposed motivates a view of our epistemic entitlements that falls well short of traditional realist commitments.

Philosophy of Science as Constructive Interpretation

Suppose now that we have (1) abandoned the idea that there are distinctively philosophical methods of investigation that are independent from those of ordinary empirical inquiry, (2) accepted that realism and instrumentalism are not best understood as distinctive philosophical views of theories or theoretical knowledge as such, but instead as cognitive attitudes that realists and instrumentalists alike take toward some particular theories or bodies of information about the world (e.g., Newtonian mechanics) and not others (e.g., Quine's hypothesis of the bodies of common sense), and (3) recognized that thoroughgoing naturalists will see the correct attitude to take toward any particular scientific theory as depending not only upon individualized consideration of the empirical achievements and further characteristics of that theory, but also on the historical record of our successive efforts to entheorize the world more generally, and perhaps further empirical evidence concerning ourselves as theorizers besides. By this point, it might seem that philosophical activity has simply been excised from, rather than integrated into, the resulting naturalistic conception of the world and our own place in it. I think the idea that embracing integrative

naturalism leaves philosophy or philosophers with nothing to contribute to the resulting project of inquiry is profoundly misguided, and this is neither simply because scientists themselves do not typically attend to what I have been suggesting is the full range of evidence relevant to deciding what cognitive attitude to adopt toward any particular theory, nor simply because there are many important questions that belong to no single scientific specialty (cf. Maddy 2007, 115ff.). I therefore wish to conclude by describing what seems to me a crucial aspect of an integrated naturalistic conception of the world and our own place within it to which I think philosophers of science are particularly well-positioned to contribute, and indeed have been contributing for quite some time, whether they realize it or not.

Let me begin with a long-standing tension in the self-conception of philosophers of science concerning whether ours is fundamentally a prescriptive or a descriptive enterprise. Many philosophers of science, I suspect, find themselves in an uneasy tension between competing impulses on this question, thinking that the job of the philosophy of science is to understand and describe the workings of science as we actually find it, but also that there is an important normative dimension to this enterprise and that it is sometimes appropriate for philosophers of science to draw conclusions from their own work concerning how scientific inquiry *should be* conducted in general or *should* proceed in a particular case. I think there is room within integrative naturalism to reconcile these apparently competing impulses in an attractive way.

Both impulses have, I think, been deservedly influential across the history of the philosophy of science itself, but each has also sometimes been *too* influential. When the prescriptive impulse has exerted an outsized influence, it has produced a variety of philosophy of science we might caricature as the "Epistemic Police." The Epistemic Police suppose that the role of philosophers is to justify and protect the epistemic authority of science while correcting and/or improving the activity of scientists by telling them what the demands of rationality or of *genuinely* scientific inquiry are and when or how those demands have or have not successfully been met. This caricature captures some of the flavor of Karl Popper's falsificationism or Imre Lakatos' methodology of scientific research programs, as well as some improvident arguments in support of contemporary formal theories of confirmation, and of course some logical positivist and logical empiricist philosophers of science were sufficiently confident in this role to be willing to replace the actual history of science with "rational reconstructions" of that history in order to more clearly draw the appropriate normative morals from it.

Perhaps it was inevitable that this conception of the relationship between philosophy and science would eventually produce a backlash, but for whatever reason, more recent philosophy of science has tended to take a far more deferential attitude toward the details of actual scientific practice and the historical record of how scientific inquiry itself has actually proceeded: most contemporary philosophy of science simply does not take itself to be in the business of telling scientists how to do science. This otherwise welcome humility has, however, sometimes threatened to generate an almost purely descriptive form of inquiry, whose caricature might be called the "Anthropology of Science." Anthropologists of Science reject the presumptuous penchant of the Epistemic Police for lecturing scientists concerning how to conform their practices more closely to the demands of rationality or the requirements of genuinely scientific inquiry, and seek instead to describe the scientific enterprise as we actually find it, as well as to draw instructive morals from the resulting more detailed and accurate conception of that enterprise (or some part of it) regarding how science is able to do what it does. Much of the work of the Anthropologists of Science consists of careful

field reports concerning the actual theoretical commitments, practices, views, and/or thinking of various groups of scientists, often preceded or followed by vague and hopeful gestures toward traditional philosophical questions about the nature of scientific inquiry that might be illuminated thereby.

I think there is an available conception of the philosophy of science that recovers the attractions of the Epistemic Police and the Anthropologists of Science while allowing us to dispense with them both. I will begin describing it by borrowing the idea of a constructive interpretation from the legal philosophy of Ronald Dworkin (1986, esp. ch. 2), who himself relies to some extent on the earlier work of Stanley Cavell. Dworkin introduces constructive interpretation by contrasting it with ordinary conversational interpretation, in which we seek to recover the actual intention or mental state of some speaker who has produced an utterance. Thus, if you say, "Would you like to go to the movies this weekend?" or "Are you planning to pay for that?", my goal in interpreting you will be to identify the mental state that actually led you to produce that utterance, and the right interpretation of the utterance will be one that correctly reconstructs that mental state. Dworkin argues, however, that such conversational interpretation is not the best way to understand how we interpret phenomena like works of art or social practices (and ultimately laws), which are better seen as targets of constructive interpretation. In such constructive interpretation, the interpreter's objective is not to reproduce the actual intention or mental state of any particular speaker or social actor, but instead to find the interpretation of the work of art or social practice that best fits the object of interpretation, while simultaneously seeking to make it, in Dworkin's repeated phrase, "the best it can be" (1986, 53). That is, constructive interpretations seek to simultaneously maximize two independent desiderata, which Dworkin calls "fit" and "acceptability": a constructive interpretation is better insofar as it better rationalizes or motivates more elements of the work of art or social practice as a whole, but also insofar as the rationale, point, purpose, aim, or value it ascribes to that work of art or social practice renders it more independently appealing, interesting, useful, attractive, significant, mind-blowing, and so on (by the interpreter's own lights and/or those of any community whom she hopes to convince to adopt or share her interpretation). For the constructive interpreter, the *right* reading(s) of *All My Sons* or "All Along the Watchtower" is simply that (or those) which manages to maximize both fit to the object of interpretation and independent attractiveness to the greatest possible degree (and, of course, the appropriate tradeoff of these virtues in any particular case is itself an interpretive decision).

Whether or not such constructive interpretation is the best option for interpreting works of art or the law, Dworkin's suggestion that it is the right approach to interpreting social practices offers a useful tool for reconceiving the philosophy of science without the Epistemic Police or the Anthropologists of Science. Philosophers of science can reconcile their competing prescriptive and descriptive impulses in an attractive way, I suggest, by seeing themselves as offering candidate constructive interpretations of the social practice of science itself. Such constructive interpretations invariably begin with some preinterpretive identification of the object of interpretation: the poem on the page, the play on the stage, or the social practice (in this case, scientific practice) as we find it. In seeking to maximize fit, a good constructive interpretation of science will seek to motivate, explain, and rationalize as much of what scientists themselves *actually* do (including much of what they say) as possible. But constructive interpretations also retain the freedom to ultimately exclude some parts of the preinterpretively identified object of interpretation as mistakes,

because fit is not all that matters. At some cost to the fit of their interpretations to the practice they are interpreting, constructive interpreters of science are free to reject parts or aspects of the *actual* practice of science as inconsistent with the most attractive constructive interpretation of most of or the rest of that practice; that is, with the constructive interpretation that makes the social practice of science the best it can be. On this view, rejecting a given part or aspect of scientific practice as mistaken or even just suboptimal is similar to rejecting a part of a play or a book, or a line in a song, as contributing nothing toward or even detracting from what is otherwise the most attractive constructive interpretation of the play or song's point, purpose, aim, value, rationale, or significance, and therefore as not actually being part of the play or book or song (as it is best interpreted) in the first place. In extreme cases, a sufficiently attractive constructive interpretation may lead us to reject a substantial part of the preinterpretively identified object of interpretation, as when some critics reject the disastrous final chapters of *The Adventures of Huckleberry Finn* as inconsistent with the best constructive interpretation of the book as a whole, even at considerable cost to the fit of that constructive interpretation with the full text of the published novel itself. For such interpreters, these final chapters of the novel are a *mistake*, in just the same way that even widespread scientific practices (of fraud, or dogmatic adherence to a theory, or poor statistical methodology) can be regarded as normatively mistaken insofar as they are, though familiar features of the practice of science as we preinterpretively identify it, nonetheless inconsistent with what we see as the independently most attractive conception of the rationale, point, purpose, aim, value, or significance of that practice.

This view of the matter allows us to understand the appeal of the Epistemic Police and the Anthropologists of Science, while dispensing with both. It sees each of these traditions as excessively fixated on just one of the evaluative dimensions of constructive interpretation (acceptability and fit, respectively) and insufficiently attentive to the other. Accordingly, it rationalizes the broad lines of criticism that rightly make each of these caricatures seem unappealing: the complaint that the Epistemic Police are seeking to characterize a philosophical fantasy rather than the actual scientific enterprise and the complaint that the Anthropologists of Science settle too quickly for mere description and prematurely abandon any substantive normative appraisal of whether and how features or elements of the scientific enterprise as we (preinterpretively) find it should be revised, rejected, resisted, or reformed as that practice is extended into the future.

Beyond respecting both the prescriptive and descriptive impulses motivating the Epistemic Police and the Anthropologists of Science, however, I suggest that seeing the philosophy of science as engaged in attempts to constructively interpret science or parts of science promises to rationalize much of the rest of what philosophers of science themselves say and do, including their efforts to develop and defend philosophical "theories" of explanation, confirmation, and the like. Bayesian confirmation theory, for example, does not seem a promising description of the process by which most scientists actually make confirmational judgments, nor do most Bayesians seem to be suggesting that we simply ignore scientists' own confirmational judgments and adopt the Bayesian apparatus as a way of making such judgments *instead*. Bayesians are best seen, I think, as offering constructive interpretations that seek to rationalize as much as possible of existing scientific judgments of comparative confirmation, of evidential virtues, and so on (though retaining the freedom to reject any particular scientific judgments or instance of that practice as mistaken), while also offering the most independently attractive conception of that practice (in part, for

familiar reasons concerning Dutch books and the like[4]). I will here limit myself, however, to detailed discussion of just one further example, in which I think the attractions of seeing the philosophy of science as aimed at constructive interpretation in this way are especially conspicuous.

Bas van Fraassen famously claims that the fundamental conflict between scientific realism and his constructive empiricist alternative concerns the *aims* of science. On a realist view, "[s]cience aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true" (van Fraassen 1980, 8). By contrast, van Fraassen's constructive empiricist position holds that "[s]cience aims to give us theories which are empirically adequate, and acceptance of a theory involves as belief only that it is empirically adequate" (1980, 12), where empirical adequacy means simply that the theory's claims about observable phenomena are true. But many philosophers of science, including perhaps most influentially Gideon Rosen (1994), have expressed puzzlement regarding just what van Fraassen is asserting. Van Fraassen is quite clear that he does not mean to suggest that the pursuit of empirically adequate theories is the aim of all or most individual scientists (an interpretation that Rosen rightly rejects as turning van Fraassen's book into "a work of inept sociology" (1994, 147)), nor is he arguing that scientists *should* adopt the constructive empiricist's aim, even if they presently do not.[5] Indeed, van Fraassen explicitly distances the constructive empiricist's perplexing claims about the aim of science from any view of what the aspirations of individual scientists are or should be with an analogy to chess, in which he says the aim of the game *itself* is checkmate, even if individual players are motivated by "fame, gold, or glory" (1980, 8). Rosen nevertheless insists that the manifest gap between constructive empiricist commitments and the (explicit or implicit) intentions of most individual scientists entails that constructive empiricism is false if taken "literally" (1994, 154) and is therefore instead "best seen as a fiction about science put forward not as true, but rather as adequate to the phenomena of scientific activity" (1994, 153).

Having proposed this fictionalist reading of constructive empiricism, however, Rosen insists that van Fraassen fails to provide any reason to adopt or even care about such a view that should appeal to anyone who does not *already* share his own commitments both to the rationality of scientific practice and to a broad form of empiricism which contends that experience is our only legitimate source of knowledge about the world. The most van Fraassen can hope to show, Rosen argues, is that constructive empiricism is the most stringent form of a broadly *empiricist* commitment to science, "which at the same time renders the phenomena of scientific activity fully intelligible and ratifiable" (1994, 162); he thinks van Fraassen says nothing that should encourage anyone who is not already and independently "committed both to the rationality of science and to empiricism in the broad sense" (1994, 163) to become a constructive empiricist in the first place.

But we need not and should not see van Fraassen as simply preaching to the choir of those antecedently devoted to both empiricism and the rationality of scientific practice.

4 The rational appeal of Bayesianism is sometimes thought to be captured in the recognition that any agent who does not assign and update probabilities in accordance with the Bayesian apparatus is open to accepting a "Dutch book": a combination of wagers on which she will be guaranteed (not just likely) to lose money.

5 It is perhaps revealing that the two interpretations of van Fraassen Rosen considers and rejects as inadequate before proposing his own are that he means simply to describe the intentional state of most scientists (the Anthropology of Science) and that he means to prescribe what those intentional states should be (the Epistemology Police).

If we instead see his sermon as an attempt to constructively interpret science as a social practice, it is the need to generate a *convincing* interpretation of this sort that leads him to seek to rationalize as much of that practice as possible, rather than a prior and undefended commitment to the rationality of science itself. Similarly, we need not follow Rosen in seeing the further virtues van Fraassen claims for constructive empiricism simply as attempts to make the attractions of the view evident to anyone who already happens to share his own broad commitment to empiricism. Instead, when van Fraassen claims that constructive empiricism exposes us to fewer epistemic risks than realism (1980, 69; 1985, 246, 280–281), that it eliminates the need to see scientists as embracing beliefs about nature on the strength of comparative virtues that are pragmatic rather than epistemic (i.e., truth-tracking) in character (1980, 87–96; 1985, 264–265, 280–281), that it disdains the realist's "empty strutting and posturing [and] display of courage not under fire…in embracing additional beliefs which will *ex hypothesi* never brave a more severe test" (1985, 255), and the like, we can see such claims as attempts to display the *independent* appeal of embracing a constructive empiricist interpretation of science itself. After all, one need not be an empiricist to want to avoid unnecessary or pointless epistemic risks, to want to limit the range of influences determining our scientific beliefs to characteristics of those beliefs that plausibly track the truth, or to find empty strutting and posturing and grandiose expressions of "courage not under fire" unappealing. Though van Fraassen clearly thinks that constructive empiricism *also* more effectively rationalizes more of the practice of science itself than does the realist alternative (1980, 73–87; see also 1994, 191), we may nonetheless see claims such as "constructive empiricism…makes better sense of science, and of scientific activity, than realism does" (1980, 73) as asserting the superiority of constructive empiricism to scientific realism on grounds of independent acceptability, as well as fit. That is, I think we can and should see constructive empiricism as offering what van Fraassen (rightly or wrongly![6]) takes to be the most appealing conception of the point, purpose, value, or aim of the social practice of science (acceptability) that also best explains, rationalizes, and defends as much of the actual activity of science as possible (fit).[7] That is, it offers a characterization of the aim of science *itself* (as opposed to the aims of all or any of its individual practitioners) that is simultaneously descriptive and normative in just the way that constructive interpretations invariably seek to be.

This proposal should also make clear why constructive empiricism is neither seeking to describe nor even particularly concerned with what most individual scientists take the aim of science to be. One of the grounds on which Dworkin argues that interpreters must construct social practices in the course of interpreting them is that there simply is no canonical participant or group of participants in such a practice whose individual intentions concerning the point, purpose, aim, or value of the practice itself *could* automatically or dispositively settle the issue (1986, ch. 2). Thus, any individual's conception of the practice in which she herself participates simply represents a candidate constructive interpretation like

6 That is, while I think van Fraassen is best seen as attempting to constructively interpret science, I think a superior constructive interpretation is offered by the epistemic instrumentalism I sketched in the previous section. Moreover, I think van Fraassen is profoundly mistaken to see the need for such constructive interpretation as arising only when we step back from scientists' own "immersion" in scientific investigation to consider the epistemic status of its products (1980, 80–82). Instead, I will go on to suggest that such constructive interpretation is (and always has been) a critical part of conducting scientific inquiry itself.

7 Indeed, I am prepared to defend this as the most attractive constructive interpretation *of constructive empiricism*, no matter what view of the matter van Fraassen himself takes.

any other – and we are more likely to think it the right one as it better maximizes both fit and acceptability, making that practice "the best it can be," while simultaneously motivating or rationalizing as much of the practice as possible. Likewise, in the scientific case, there is no canonical person, group, or institution whose views on the aim of science could plausibly serve to simply settle the question authoritatively in the way that the actual intentions of a speaker do in the case of conversational interpretation. Thus, though scientists speak with especially detailed knowledge of their own practices, their own views about the point, purpose, aim, or goal of science itself do not automatically have any special standing or claim to authority. When scientists express such views, they themselves are acting as constructive interpreters of science: they propose views of what science *itself* has been aiming to do all along and of how to best extend that practice going forward in light of that aim. Such proposals must be evaluated in precisely the same ways as those offered by philosophers of science, just as empirical hypotheses proposed by philosophers should be evaluated in just the same ways as scientists' own empirical hypotheses.

These two forms of evaluation are not the same, however. For too long, too many philosophers of science have thought of themselves as engaged in the sort of hypothesis-testing with which they see scientists and scientific inquiry as primarily concerned, using details of the practice or history of science as the "data" for which any philosophical "hypothesis" about the nature of science or some part of science is responsible. But philosophy of science is not science, nor are its central activities best assimilated to hypothesis-testing. Perhaps the most important difference is that scientists are not free to normatively evaluate and reject their data in the same way that philosophers of science normatively evaluate the practices of science itself. That is, while scientists can and certainly do evaluate and reject particular data as suspect or unreliable, they are not free to reject what they take to be genuine empirical results as *normatively inappropriate* or as *mistakes* simply because they conflict with what is otherwise the most exciting, impressive, interesting, mind-blowing, or in some other way desirable theoretical account of a given natural domain that would account for the rest of the data we have regarding it. By contrast, philosophers of science are free to propose constructive interpretations of science as we (preinterpretively) find it that reject particular aspects or instances of that practice as normatively inappropriate, misguided, and/or inconsistent with the best conception of the point, purpose, goal, aim, or rationale of scientific practice as a whole. In fact, I have suggested that doing so is part of the very point of conducting philosophy of science in the first place.

To recognize this crucial difference between science and the philosophy of science is not, however, to see the latter as seeking to investigate the world or our own place within it by means of some distinctively philosophical form of inquiry that is somehow independent of ordinary empirical investigation itself. It is instead to suggest that forming a conception of *what we ourselves are doing* when we use those empirical methods to investigate the world and our own place within it is *already* part and parcel of any such integrated naturalistic inquiry. Indeed, at least since the origin of modern science in the earlier intellectual traditions of natural philosophy and natural history, articulating such conceptions and contesting their adequacy has been a central concern of scientists themselves. The constructive interpretation of science is thus something that scientists themselves already do (and have always done) in trying to situate their own practices within a broader vision of the world and our own place within it. I suggest that any satisfying conception of our "inherited world theory as a going concern" will have to include not only our best scientific theories but also our best constructive interpretation of the point, purpose, value, or significance of

constructing, exploring, and testing such theories, including what kind(s) of knowledge of the world we think they can and cannot give us under various kinds of epistemic circumstances.

Of course, the suggestion that philosophy of science is in the business of the constructive interpretation of science is neither purely a description of everything that philosophers of science presently do (still less what they already think of themselves as doing), nor purely a recommendation regarding what I think they ought to be doing instead, because philosophy of science is itself a social practice, and therefore itself stands in need of constructive interpretation, which is to say, a simultaneously descriptive and prescriptive conception of its aim, purpose, value, point, goal, or significance and of how we might accordingly continue that collective practice into the future. By its own lights, the constructive interpretation *of the philosophy of science* I have just proposed will have to compete with others on the grounds of both fit and acceptability, and I suggest that this interpretation not only best fits or rationalizes what philosophers of science say and do but also offers an independently attractive conception of what we have been up to all along. When it comes to interpreting ourselves, it's turtles all the way down.

Acknowledgments

Thanks to my LPS colleagues Jeff Barrett, Jim Weatherall, and especially Penelope Maddy for many useful conversations regarding this material. Thanks also to generous audiences at the Universidad Autónoma Metropolitana in Mexico City and at a conference held at Lingnan University in Hong Kong. I am grateful to the University of Pittsburgh's Center for the Philosophy of Science, where some work on this chapter was completed while I was Senior Fellow there, and to the Australian National University, where I received helpful feedback on it as a Visiting Fellow.

References

Barrett, J.A. (2003). Are Our Best Physical Theories Probably and/or Approximately True? *Philosophy of Science* 70: 1206–1218.

Blackburn, S. (2002). Realism: Deconstructing the Debate. *Ratio* 15: 111–133.

Dworkin, R. (1986). *Law's Empire*. Cambridge, MA: Belknap.

Ladyman, J. and Ross, D. (with Spurrett, D. and Collier, J.) (2007). *Every Thing Must Go. Metaphysics Naturalised*. Oxford: Oxford University Press.

Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science* 48: 19–49.

Maddy, P. (2007). *Second Philosophy: A Naturalistic Method*. Oxford: Oxford University Press.

Quine, W.V.O. (1951/1980). Two Dogmas of Empiricism. In *From a Logical Point of View*, 2nd edn. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1975/1981). Five Milestones of Empiricism. In *Theories and Things*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1976). Posits and Reality. In *The Ways of Paradox and Other Essays*, 2nd edn. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1981). *Theories and Things*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1990). *Pursuit of Truth*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1995). *From Stimulus to Science*. Cambridge, MA: Harvard University Press.

Rosen, G. (1994). What is Constructive Empiricism? *Philosophical Studies* 74: 143–178.

Rosenberg, A. (2014). Disenchanted Naturalism. In *Contemporary Philosophical Naturalism and its Implications*, edited by Bana Bashour and Hans D. Muller, pp. 17–36. New York: Routledge.

Stanford, P.K. (2005). Instrumentalism. In *The Philosophy of Science: An Encyclopedia*, edited by Sahotra Sarkar and Jessica Pfeifer, pp. 400–405. New York: Routledge.

Stanford, P.K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.

Stein, H. (1989). Yes, But… Some Skeptical Remarks on Realism and Anti-Realism. *Dialectica* 43: 47–65.

van Fraassen, B. (1980). *The Scientific Image*. Oxford: Clarendon.

van Fraassen, B. (1985). Empiricism in the Philosophy of Science. In *Images of Science: Essays on Realism and Empiricism*, edited by Paul Churchland and Clifford Hooker, pp. 246–308. Chicago: University of Chicago Press.

van Fraassen, B. (1994). Gideon Rosen on Constructive Empiricism. *Philosophical Studies* 74: 179–192.

8

"The Horrid Doubt": Naturalism and Evolutionary Biology

VALERIE GRAY HARDCASTLE

> But then with me the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey's mind, if there are any convictions in such a mind?
>
> Charles Darwin, to William Graham (July 3, 1881)

We can divide naturalism into two flavors: ontological and methodological. In brief, ontological naturalists hold that there are no supernatural entities, processes, or events in the universe. Everything is or reduces to something physical, where "physical" is understood in the way that science understands physics. Put more colloquially: there is no "spooky stuff" in our world. Methodological naturalists hold that science and the concomitant scientific method is the best and only way to investigate the nature of the furniture of our universe. Science investigates physical objects and interactions, and, in so doing, pays attention only to physical things. Anything supernatural is simply beyond the reach of science. One can, of course, be both types of naturalist at the same time, and most naturalists are just that.

Wanting to explain how things make a causal difference is an underlying motivation for both types of naturalism. Consequently, naturalists of both stripes believe that only earthly-type causes are the difference-makers in our world. In particular, evolutionary biologists who are also naturalists believe that natural selection is a purely causal interaction that follows basic rules discovered and understood by science. The evolution of humans from other organisms has a purely natural story that is really no different in kind from any other scientific account of the appearance and evolution of any other entity or process.

I should note that it could, in theory, be the case that non-natural entities causally interact with natural ones to produce an object, like, perhaps, the hand of God creating humans. There are no *a priori* considerations that rule out this possibility. However, naturalists deny such claims on *a posteriori* grounds: we do not need to postulate the existence of a supernatural being to explain our own existence; therefore, by Occam's razor or similar considerations, we shall not. Contemporary science has adopted the theoretical constraint

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

that only physical interactions can lead to physical effects (more on this further on). Supernatural events, processes, and objects are simply outside the realm of science, and, therefore, outside the realm of naturalism.

While most of the world has been content to let science as a whole trundle along with its naturalistic assumptions without substantial opposition (though see Chapters 14 and 15), evolutionary biology has been singled out as perhaps an exception to that rule. Many believe that the evolution of humans requires the existence of something supernatural. For example, a recent poll presented at the American Association for the Advancement of Science (AAAS) (Ecklund and Scheitle 2014) indicated that, while most people understand that evolutionary theory claims that humans evolved from a primate-type ancestor, many deny that the theory is true. Almost 40% of the 10 000 respondents thought science "should be open to considering miracles in their theories and their explanations." Indeed, as reported in *Science* (Miller, Scott, and Okamoto 2006), a previous survey of citizens from the United States and over 30 European countries revealed that only adults in Turkey expressed more doubts about evolutionary biology than those from the United States.

This chapter takes a brief look at the history of the Western conflict between evolutionary biology and Christianity, and then examines a contemporary argument for why hypothesizing the creation of humans without also positing the existence of some supernatural being is a mistake. It closes by recounting the most recent skirmishes in the United States regarding how evolution should be articulated and understood in the public arena: Is it scientific or not?

Early Conflicts, 1692–1872

In contrast to what some may think, the concept of biological change over time had been a point of conversation among religious leaders and scientists well before Charles Darwin's theories of evolutionary biology were popularized. In the late 1600s, for example, the Christian priest and botanist John Ray asked, "why such different species should not only mingle together, but also generate an animal, and yet that that hybridous production should not again generate, and so a new race be carried on" (1692). The 18th-century Christian and scientist Carl Linnaeus pondered the evolution of plant species toward the end of his life:

> We imagine that the Creator at the actual time of creation made only one single species for each natural order of plants, this species being different in habit and fructification from all the rest. That he made these mutually fertile, whence out of their progeny, fructification having been somewhat changed, Genera of natural classes have arisen as many in number as the different parents, and since this is not carried further, we regard this also as having been done by His Omnipotent hand directly in the beginning; thus all Genera were primeval and constituted a single Species. That as many Genera having arisen as there were individuals in the beginning, these plants in course of time became fertilized by others of different sort and thus arose Species until so many were produced as now exist…these Species were sometimes fertilized out of congeners, that is other Species of the same Genus, whence have arisen Varieties. (1762, as quoted in Wilkins 2009, 72)

He had initially embraced Aristotle's notion that species were immutable, but later rejected that idea, even though he still believed that God was the one who set something like evolution in motion, through His "special creation" of all the original plants and animals on Earth.

In the mid-1800s, the Augustinian friar Johann Gregor Mendel aligned himself with Linnaeus' ideas when he concluded that new species of plants could indeed arise under certain limited conditions. At the time, his work was seen to be more about hybridization than evolution, but it is clear that Mendel himself thought that change in species was possible. Unlike his contemporaries (including Darwin), Mendel believed – and his experiments bore out – that some traits are passed on whole, instead of being blended with other traits of the same type. Crossing a tall plant with a short plant produces another tall plant, instead of a plant that is of middling height. If these so-called dominant traits are passed on enough of the time, then the entire species can change into something else.

At roughly the same time these discussions were happening, the French anatomist and zoologist George Cuvier (1769–1832) developed a theory of extinction through his comparison of fossils with currently living organisms, founding modern paleontology. In contrast to his contemporary Jean-Baptiste Lamarck, Cuvier did not believe that species themselves evolved. Instead, they were invented anew periodically through catastrophic events, a hypothesis known as "catastrophism." British followers of his work, especially the geologists William Buckland and Robert Jameson, interpreted Cuvier's suggestion of a catastrophic event as the Flood described in the Old Testament. In addition, the newer species were better adapted to their surroundings, which they took as a sign of God's hand in the changes.[1]

Cuvier's critics, like Lamarck, Étienne Geoffroy Saint-Hilaire, Robert Grant, and Robert Chambers, maintained that biological change happened gradually over time, though most also saw this as evidence for a benevolent God. A Divine Creator had a plan for evolution that was ultimately to culminate in humankind. Sir Charles Lyell (1797–1875) also saw adaptation to changing environments as a sign of a supernatural Creator, but he argued for a uniform pattern of extinction and replacement over time.

In all of these cases, though, scholars were looking for law-like regularities and not divine intervention in the unfolding speciation. Even though these theorists were deeply religious, they believed that God's interventions and interactions with the physical world were quite minimal. God's plan, once set in motion, was to unfold without miracles. Once God got things going, He rapidly moved to the background, and the world could unfold unproblematically according to the rules He created. These rules and regularities are what science uncovers when it investigates the world, and they do not require ongoing supernatural intervention to maintain themselves.

It is against this background that Darwin published *On the Origin of Species* in 1859. Many Christians of that time accepted evolution insofar as it could be reconciled with God designing the universe in some fashion. The Church of England priest Charles Kingsley (1819–1875), a personal friend of Darwin's, accepted that creation is an indirect process controlled by divine law. He wrote to Darwin in response to receiving an advance copy of his book that he had "long since, from watching the crossing of domesticated animals and

1 Today, scientists explain abrupt changes in the fossil record with the theory of "punctuated equilibrium." Punctuated equilibrium holds that most species exhibit stasis most of the time. Only occasionally do species branch into new species, in a process called cladogenesis, and this happens very quickly (in geologic time), not gradually. Niles Eldredge and Stephen Jay Gould advanced these ideas in a landmark essay in which they claimed that the fossil record shows that Darwin's hypothesized gradual morphing of one species into another is essentially nonexistent, and that stasis dominates what we can see through the fossil record (Eldredge and Gould 1972).

plants, learnt to disbelieve the dogma of the permanence of species" (1859, as quoted in Darwin 1887). The Archbishop of Canterbury, Frederick Temple, also believed that evolution was compatible with theology. In his Eight Brampton Lectures on the Relations between Religion and Science, Temple states unequivocally that the "doctrine of Evolution is in no sense whatever antagonistic to the teachings of Religion" (Temple 1884). Darwin himself, in his second edition of the *Origin*, adds the phrase "by the Creator" to the final sentence of his conclusion:

> I believe that animals have descended from at most only four or five progenitors, and plants from an equal or lesser number. Analogy would lead me one step further, namely, to the belief that all animals and plants have descended from some one prototype. But analogy may be a deceitful guide. Nevertheless all living things have much in common, in their chemical composition, their germinal vesicles, their cellular structure, and their laws of growth and reproduction. We see this even in so trifling a circumstance as that the same poison often similarly affects plants and animals; or that the poison secreted by the gall-fly produces monstrous growths on the wild rose or oak-tree. I should infer from analogy that probably all the organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed by the Creator. (1860, 428)

But this position of minimal Godly intervention was controversial. From the first publication date, significant opposition existed to the idea of natural selection, and many sought a more purposeful mechanism behind creation. Sir John Herschel asserted that "an intelligence, guided by a purpose, must be continually in action to bias the direction of the steps of change – to regulate their amount – to limit their divergence – and to continue them in a definite course" (1861, 12). The Scottish scientist Sir David Brewster rejected many Darwinian ideas as being antireligious in his article "The Facts and Fancies of Mr. Darwin," and further held that "every part of the human frame had been fashioned by the Divine hand and exhibited the most 5 piphenom and beneficent adaptions for the use of men" (1862, 171).

By the late 1860s, however, these tensions among scientists and theologians resolved themselves into what became known as theistic evolutionism. The general public also supported this view, by and large. Richard Owen (1860) supported a theory of derivation in which species innately changed over time to increase variety and beauty, thus illustrating divine creative purpose. Both Owen and St. George Jackson Mivart held that purely natural selection could not explain the patterns and wide variation they discerned in the world around them; it had to come from something divine. Similarly, in 1867, George Campbell, the Eighth Duke of Argyll, published *The Reign of Law*, which explained the completely nonfunctional beauty in birds' plumage as something the Creator did by putting together the laws of nature for the amusement and delight of humans. He tried to bring design and evolution together, though, by hypothesizing that the Creator's laws of variation also crafted some rudimentary organs that would fill a future need. In the main, though, both those in religious quarters and those in the biological attacked these seemingly ad hoc hypotheses that attempted to bridge theology and natural selection.

However, Darwin's (1872) speculation that humans may have "descended from a hairy quadruped, furnished with a tail and pointed ears" promptly gave rise to more opposition to his theory from religious leaders. Cardinal John Henry Newman wrote in 1868: "Mr. Darwin's theory need not then to be atheistical, be it true or not; it may simply be suggesting a larger

idea of Divine Prescience and Skill…and I do not [see] that 'the accidental evolution of organic beings' is inconsistent with divine design – It is accidental to us, not to God" (1868/1971). In other words, even though we may think that evolution unfolds without apparent design, we are not wise enough to perceive God's divine purpose in evolution.

This suggestion for how to reconcile theology with natural selection and the theory of evolution lives on today; in fact, it is the most accepted view of how to understand the origins of humans in the Western world. We limited humans can discern natural regularities in the world around us, and it is upon these rules that we base our sciences. However, there remains a supernatural presence out there to which we are not privy. This Presence made the rules we discover; in particular, the rules that God established led to our evolution from other species over time. This sort of compatibilistic view is a type of methodological naturalism – we must constrain our scientific investigations to the natural world, but we can also believe that there is (or could be) more to the universe than what we can scientifically divine.

Contemporary Conflicts

Not all Christians, however, accept this possibility. Some branches of Christianity interpret (or try to interpret) the first two books of Genesis quite literally, which entails that their religious tenets are fundamentally incompatible with any evolutionary account of the origin of our species. The earth is only a few thousand years old on this account, in which case, among other things, there simply has not been enough time for evolution by natural selection to have occurred. God made the world a relatively short time ago; as part of the world, he also created the fossils and rock strata that we now measure and date using our methodologically natural assumptions and our best science. Our science tells us that the world and the creatures on it are millions of years old, which presumably, God intended us to find. Our reason is not such that we are able to understand why God would do these things, but He did them nonetheless.

At the same time, many committed to the truth of Darwinian-style evolution also believe that there is a fundamental incompatibility between orthodox Christianity and the tenets of natural selection (see, e.g., Dawkins 1986; 2003; Dennett 1995). Evolution by natural selection is both purposeless and blind; it does not and cannot have a hidden meaning or reason. In particular, it cannot be something that would necessarily culminate in the evolution of humanity, as some Christians believe. As Dawkins writes:

> All appearances to the contrary, the only watchmaker in nature is the blind forces of physics, albeit deployed in a very special way. A true watchmaker has foresight: he designs his cogs and springs, and plans their interconnections, with a future purpose in his mind's eye. Natural selection, the blind, unconscious automatic process which Darwin discovered, and which we now know is the explanation for the existence and apparently purposeful form of all life, has no purpose in mind. It has no mind and no mind's eye. It does not plan for the future. It has no vision, no foresight, no sight at all. If it can be said to play the role of watchmaker in nature, it is the *blind* watchmaker. (1986, 5)

If evolutionary theory (some day, when all is said and done) can explain and predict (or retrodict) all evolutionary phenomena, then there simply is no role for a divine influence to play.

In both perspectives, the conflict stems, on the one hand, from the Christian's conviction that the creatures on this earth were designed and, on the other, from the Darwinian's view that all organisms were produced by a blind, random, and unguided process in natural selection. In other words, the conflict comes from a disagreement over whether ontological naturalism is true. While both sides can (in theory, at least) agree that science presumes naturalism as a methodology, Christians hold that a supernatural God is responsible for everything, while the followers of a Dawkinsian approach presume a much more sparse ontology. No God is needed to make our evolutionary theories function; therefore, no God exists.

Not all evolutionary biologists subscribe to Dawkins' perspective, however. They instead point out that there is nothing in the theories of natural selection that claims that changes across time are random in the sense of being uncaused. Instead, the mutations associated with evolutionary theory are random in the sense that they do not necessarily enhance a creature's reproductive fitness. Thus, Ernst Mayr writes: "When it is said that mutation or variation is random, the statement simply means that there is no correlation between the production of new genotypes and the adaptational needs of an organism in the given environment" (1988, 99). In this case, God could have orchestrated the entire process; He could also be overseeing it in minute detail. God could be causing all the random mutations that natural selection operates over. We just will not be able to discover whether this is the case because the putative randomness of evolutionary mutations prevents us from uncovering any larger pattern to the changes.

Peter van Inwagen (2003) argues that the claim that evolution demonstrates that human beings and other living creatures have not been designed is neither part of nor a consequence of any scientific theory, but is itself a metaphysical or theological addition to science. Science must be agnostic about entities, events, and processes that it cannot measure or investigate. Therefore, those who presume more than what science can show are presuming a metaphysics not justified by scientific investigation. Both the Christian fundamentalists and the ontological naturalists are guilty of this charge. (For more on this idea, see Fetzer 2007; 2011.)

It is here that the division between ontological and methodological naturalism with respect to evolutionary theory shows itself most clearly. Ontological naturalists would have to side with Dawkins and Dennett: there is no God and we do not need to posit one to explain any physical event in the universe. Methodological naturalists could be more tentative. They could side with van Inwagen: the methods of science tell us how evolution occurs and explain all the interactions we need to account for natural selection. However, methodological naturalists would have to remain silent on the question of the existence of God, since the question itself is about things not in this world. The differences a God would make are not empirically detectible; therefore, about this issue, methodological naturalists can remain silent.

For many, however, especially Christians, this is not simply an academic issue regarding how one puts together one's metaphysical commitments. A universe without God also means a universe devoid of meaning and morality. As the theologian David Ray Griffin so eloquently explains, "Insofar as Neo-Darwinists believe that to accept evolution is to accept religious-moral implications, they are claiming that we cannot have evolution without tears, that is, without accepting the conclusion that human lives, including all our finest achievements – such as the discovery of the evolutionary nature of our universe – are ultimately meaningless" (Griffin 2006, 13, as cited in Fetzer 2011, 392). If the laws of nature are

arbitrary and meaningless, then, according to this perspective, the universe itself is arbitrary and meaningless. And without meaning, one cannot have morality. (To unpack exactly how this argument runs would take us beyond the scope of this chapter; for further amplification, see Griffin 2006 and Fetzer 2011).

Ontological naturalists are not without a response; James H. Fetzer notes that if "we are the entities that impart meaning to the universe and through the exercise of our critical reasoning produce an understanding of morality that, ironically, transcends our biology, is that not to the greater glory of man, if not of God?" (2011, 393). One may not need a God for meaning or morality, though these may be things outside the purview of science. Humans have other ways of understanding and appreciating the world around them beside science. Meaning and morality could be things that we create and then impute to the world around us. This would not give us transcendent and eternal meanings and moral principles, so ultimately this approach would likely not satisfy those who desire such things, but it does give ontological naturalists a way to understand meaning and morality in a purely physical world.

It is difficult to see which side has the upper hand in this dispute, because both sides are starting from positions so removed from each other, and neither can agree on common points of reference. From the perspective of a methodological naturalist, one who shies away from any sort of deep metaphysical commitments, the dispute is essentially a draw, for both sides are helping themselves to metaphysical assumptions that empirical evidence simply can neither justify nor support.

The Evolutionary Argument against Naturalism

There is a second type of argument against ontological naturalism in evolutionary theory. Known as the "evolutionary argument against naturalism," or EAAN, it claims that believing both evolutionary theory and ontological naturalism is self-defeating, in that if both are true, then we humans cannot know that both are true. In short, if evolutionary theory and ontological naturalism both accurately describe the world, then we can never know that fact. Naturalism undercuts its own justification.

Proposed most recently by the Christian philosopher Alvin Plantinga (1993; 2000; Plantinga and Tooley 2008), EAAN has its roots in the work of Arthur Balfour (1906) and C.S. Lewis (1947), among others (including Richard Taylor, Stephen Clark, Richard Purtill, J.P. Moreland, Victor Reppert, James Jordan, and William Hasker). We summarize Plantinga's 1993 version here.

Let us assume that N is ontological naturalism, which Plantinga defines as the belief that there is no such thing as God or anything resembling God; E is the belief that evolutionary theories are true and humans evolved just as the theory asserts; and R is the proposition that our cognitive faculties are "reliable," in the sense that the vast majority of their outputs are true. If so, the probability of R given N and E, that is, $P(R|N\&E)$, is low or inscrutable. Why is this?

Plantinga starts with the assumption that unguided natural selection would not select for any sort of truth-creating mechanism in humans; instead, it would only select for behaviors that aid in reproduction or survival. If epiphenomenalism is true, and beliefs do not cause behaviors (or the semantic contents of beliefs do not cause behaviors), then beliefs will be "invisible" to evolution, and hence $P(R|N\&E)$ will be low or unknowable. On the

other hand, if beliefs do cause behaviors, then either beliefs are maladaptive, in which case P(R|N&E) will be low, since R will be selected against, or beliefs are adaptive. However, even if they are adaptive, they may still be false. Natural selection would have no reason for selecting true but nonadaptive beliefs over false but adaptive beliefs. Thus, P(R|N&E) could still be low. Therefore, Plantinga concludes, to assert that naturalistic evolution is true also asserts that one has a low or unknown probability of being right. And this conclusion, he thinks, means that we cannot hold that a belief in naturalistic evolution is true because ascribing truth to both naturalism and evolution is inconsistent.

Michael Ruse reminds us that Plantinga's arguments are only against ontological naturalism and not methodological naturalism. So, at best, Plantinga is telling us that we cannot know "the world in some absolute sense," or "metaphysical reality" (Ruse 2004, 188). That does not seem to be a very strong argument against evolutionary theory, or any science, from his perspective, as science does not claim to tell us about the world as it actually is in any deep sense. Science is really about predicting and controlling the world around us, the world we can access and interact with. If there are other things out there that we cannot touch, then they lie beyond the reach of scientific hypotheses. This, of course, is the perspective of a methodological naturalist, not an ontological one.

Another way of looking at the same response is to note that, for Plantinga's argument to hold sway, one must believe that having the Truth "with a capital T" is not only important, but is a product of science. That is, the scientific method must give us Truth about our universe, and not just explanation, prediction, understanding, and control. Otherwise, claiming that the reliability of our beliefs is inscrutable has no argumentative bite. This is quite a strong realist position. Plantinga assumes without question that we have access to Truth, and that believing in naturalistic evolution would deny us this access, or at least would illustrate how this access might be illusory. One could adopt a less stringent realist position (or an antirealist one) and agree that such access to Truth is a phantom.

Following the seminal work of Amos Tversky and Daniel Kahneman in the 1970s, psychology has developed a cottage industry devoted to highlighting that at least some of our cognitive faculties are unreliable. To take an example from Tversky and Kahenman's original work, consider the case of a hypothetical "Steve" who "is very shy and withdrawn, invariably helpful, but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail" (Tversky and Kahneman 1974, 1124). When people are asked to assess Steve's most likely profession from a list of possibilities (farmer, salesman, pilot, librarian, engineer, etc.), they invariably select the profession (librarian) that fits best with their stereotype or mental image of what members of that profession are like, without paying any attention to things like base rates. That there are many more farmers, salesmen, and engineers than there are librarians (and that most librarians are women, while farmers and engineers are overwhelmingly men) does not enter into their assessments. We judge based on representativeness and similarity, not on actual probabilities. This sort of research tells us that the way we form beliefs about the world systematically results in errors in reasoning. We cannot help but have false beliefs about the world.

However, if this is correct, then one might conclude that evolution and naturalism are actually better at explaining our imperfections than is an all-powerful and all-knowing Creator. In other words, if we are indeed fallible, and we possess false beliefs, then that is stronger evidence for some "blind" process, like natural selection, than for a perfect God having brought us into being. That (some of) our beliefs are not true is a consequence of accepting science – not only evolutionary theory, but our best psychological theories as well.

Similar arguments are made from our biology, as well. Stephen Jay Gould (1980) argues, for example, that the panda's thumb (which is not a thumb at all, but an extension of the radial sesamoid) is poorly designed. Gould believes that such instances support the existence of blind evolution over an all-knowing divine designer. We can see similar examples of poor design in us humans, too, which seems to be evidence against those who believe that evolution was designed to culminate in humanity. For example, we all have a tiny spot known as the optic disk in the back of our eyes, where the optic nerve and approximately 1 million ganglion cells exit the eyeball. Because no rod or cone cells can be located there, we each have a blind spot in that part of the retina, a very small gap in our visual field.

Researchers have proposed a number of different explanations for why we do not notice this blind spot. Most believe that the opposite eye compensates for the missing visual information. In other words, the brain fills in the missing information as best it can using other visual cues in the environment. Instead of perceiving a gap in our visual field, we create a full yet false perception (which, of course, would mean that we constantly have false beliefs about what we are seeing in the world around us).

Here we have two instances of poor design. First, there is no reason that the optic nerve and the ganglion cells have to exit the eye in the middle of the retina. In theory, they could just as easily have exited in another location, one that would not create a blind spot. Second, having the brain make up information to fill in the blind spot means that we are not aware that we are not actually seeing the entire visual area in front of us. We therefore do not shift our eyes or our head to compensate for the blind spot, even when doing so might be to our advantage. Again, some take examples such as these as evidence that evolution is blind: no good designer would do such poor work.

We could say that our beliefs, the panda's thumb, and our visual system "satisfice." That is, they are good enough to get the job of successful living done most of the time, but they are not ideal; they could have been done better. But our cognitive faculties, by and large, satisfy the needs of their possessors. They help us survive and reproduce enough to get our genes into the next generation.

Returning to the issue that Plantinga raised, if our bodies and our minds "satisfice," then that means that our beliefs about the world are false a lot of the time. For those in search of absolute truth and certainty, this can be a bleak conclusion. It would also be a bleak conclusion for those who maintain that they have access to the absolute truth of God's word and intentions. And it is likewise a bleak conclusion for those who claim the all-too-human enterprise of science is an engine of truth.

Impact of the Debates

Perhaps unlike other debates regarding naturalism, the question of how and whether God fits with evolution has recurring cultural, political, and educational impacts. We can see this most publicly in the United States, where some Christians loudly dispute the paleontological evidence for the evolution of humans from other species by natural selection. These are the Christians who argue for a more literal interpretation of the Bible, sometimes claiming that it too is scientific. This debate plays out in political discussions across the country regarding what should be included in the public school curriculum: evolution or creationism. Should both views be taught? Should only one? If one, which one? If both, then should they both be taught as science? These discussions have broad implications not only for our

notions of what counts as science, but also for science education, free speech, the separation of church and state, and the nature, power, and role of theology in communities.

Over the past 15 years, the Kansas State Board of Education has been a microcosm of what is happening across the country. In 1999, the board adopted a set of curricular standards for science classes that its local school districts had to follow, which presented evolution, the estimated age of the earth, and the Big Bang as controversial hypotheses. Not surprisingly, this move was roundly criticized, and citizens voted in a new board in 2001, which promptly restored the science curriculum recommendations to what they had been previously. Subsequently, the composition of the board shifted again, and in 2005, the board again voted to adopt a set of science standards in which evolution was treated as a controversial hypothesis. As part of the revision, the Kansas State Board of Education also changed its definition of science so that it could appeal to supernatural forces if the evidence warranted it. Many believe that the board did this so that teachers would be free to teach creationism in their schools. This time, the move was publicly criticized by the American Association for the Advancement of Science (AAAS), the American Institute for Biological Sciences (AIBS), the Kansas Association of Teachers of Science (KATS), the National Academy of Sciences (NAS), the National Science Teachers Association (NSTA), a group of 38 Nobel laureates, and the committee of scientists and teachers who wrote the set of standards before the board decided to revise them (cf. discussion in Branch 2011). Then, once again, in 2007, the board's members changed and it revised the standards back to what they originally were.

What has happened in Kansas is not isolated by any means. In 2005, the National Center for Science Education noted over 80 disputes concerning the teaching of evolution in schools (Branch 2006). In the same year, the NSTA reported that almost one-third of its members claimed that they had been pressured to omit or downplay evolution in their science classes, as well as to include nonscientific alternatives to evolution in their curricula (NSTA 2005). In that same year, the constitutionality of teaching a version of creationism known as "intelligent design" in public schools was challenged in Dover, Pennsylvania, which had adopted a policy that required teachers to tell their students that evolution was not a fact and that intelligent design was a scientifically credible alternative perspective. Judge Jones ruled in *Kitzmiller v. Dover Area School District* that intelligent design "is not science and cannot be adjudged a valid, accepted scientific theory as it has failed to publish in peer-reviewed journals, engage in research and testing, and gain acceptance in the scientific community."

This ruling was important because it went beyond adjudicating whether teaching intelligent design in school was constitutional: it concluded that intelligent design is not scientific. The question of how to demarcate science is one philosophers have been debating in earnest since the era of logical positivism in the late 1920s, when they concerned themselves with carving off science from other types of knowledge. Judge Jones offered a pragmatic reading of science: science is what scientists do. Philosophers have tried to draw more principled distinctions, based on notions such as verifiability, falsifiability, testability, and so on. The demarcation controversy is relevant to the discussion of naturalism and evolution, in that it analyzes whether science is ontologically or methodologically naturalistic, or whether its assumptions and methods might allow for supernatural events if the evidence demanded such.

Philosophers widely agree that, whatever science is, intelligent design is not science. Sahotra Sarkar (2007; 2011; see also Sober 2004), for example, argues that even the mildest

versions of intelligent design, which state that some intelligent designer or other created the universe with its natural laws,[2] provide no rules for what counts as "design" or "intelligent," and that any definitions or descriptions we come up with are only by analogy to what we know from human design techniques. Moreover, intelligent design faces a dilemma. If a primordial intelligent designer is subject to physical laws (but there is no evidence for a designer's existence), intelligent design is not a science because of the lack of evidence for the designer's existence. If the primordial intelligent designer is not subject to physical laws (but we know nothing about designers outside of physical laws, to be able to make any claims about how they design things), intelligent design is not a science because it involves something beyond scientific investigation.

Not all philosophers agree with this analysis. William Dembski argues that "the world contains events, objects and structures that exhaust the explanatory resources of natural causes and that can be adequately explained only by recourse to intelligent causes…Precisely because of what we know about natural causes and their limitations, science is now in a position to demonstrate intelligent causation rigorously" (1999, 47). In particular, we find design if there is sufficient "complex specified information (CSI)" (Dembski 2002). Low-probability events have a high amount of CSI, and they show design if they can be viewed as conforming to some pattern. However, many have pointed out, Dembski fails to define his terms; in particular, how one is to understand "information" and "pattern" is unclear. (Shallit and Elsberry 2004 argue that Dembski's approach is in fact incoherent.) More importantly, though, if we could work out what CSI was, it still would not present positive evidence for intelligent design *per se*, but rather would be a challenge for evolution to address, which most scientists believe it has already (see discussion in Pennock 2011 for more details).

Another tack taken in arguing that intelligent design is a valid scientific approach to understanding evolutionary change over time is Robert O'Connor's notion of "local design." Some particular event or process can only be accounted for by hypothesizing divine intervention:

> If one can establish on the basis of the relevant empirical laws and natural conditions, that some phenomenon, E, is highly improbable, and if one can further establish that E fulfills a sufficiently narrowly constrained, yet necessary function within those causal processes, one will have successfully established that E is the result of neither regularity, chance, nor their combination. Even though E is as it must be in order to render a particular natural process functional, it does not result from any known natural laws and conditions, nor is it within the limits of chance. The exceedingly low probability of E reflects its extraordinary complexity, and the narrow range of functionality reflects its specificity. (O'Connor 2003, 70)

As Shanks and Green (2011) point out, this type of argument rejects not only ontological naturalism, but methodological naturalism as well. Indeed, proponents of this argument view methodological naturalism as an arbitrary constraint on the kinds of explanation

2 The biochemist Michael Behe articulates one of the mildest versions: "The core claim of intelligent design theory is quite limited. It says nothing directly about how biological design was produced, who the designer was, whether there has been common descent, or other such questions. Those can be addressed separately. It says only that design can be empirically detected in observable features of physical systems" (2001, 15). He suggests that those who want to avoid theological implications might consider the design they see in the world around them to be the result of space aliens or time-travelers or some such.

science can offer. However, theists can easily avoid the dilemma O'Connor purports to have identified. They need not claim that God's actions are incompatible with efficient causal explanations based on the laws of nature and antecedent conditions: just as many have argued since before the time of Darwin that God's actions can be construed as initiating the laws of nature. In this case, the idea of divine creation is compatible with methodological naturalism. But it is not then a scientific claim.

Such appears to be the consensus of the scientific communities. The NAS has opined that, "Creationism, intelligent design, and other claims of supernatural intervention in the origin of life or of species are not science because they are not testable by the methods of science" (NAS 1999, 25). The NSTA agrees:

> Although no single universal step-by-step scientific method captures the complexity of doing science, a number of shared values and perspectives characterize a scientific approach to understanding nature. Among these are a demand for naturalistic explanations supported by empirical evidence that are, at least in principle, testable against the natural world. Other shared elements include observations, rational argument, inference, skepticism, peer review and replicability of work…Science, by definition, is limited to naturalistic methods and explanations and, as such, is precluded from using supernatural elements in the production of scientific knowledge. (2000)

As part of his preparation for testifying at the Kitzmiller trail, the philosopher Robert Pennock searched the major indices of scientific journals to uncover any appeal to the supernatural in scientific studies. He reported that after going through databases covering tens of thousands of peer-reviewed scientific articles, he found only a few that referenced the supernatural, and those were mostly medical studies concerning how to manage patients' beliefs in the supernatural (Pennock 2005). He found only one article that took supernatural possibilities seriously (Levin 1996): an article on the power of prayer as an alternative to or supplement to medicine. But even that author, Pennock reports, explicitly acknowledged that considering such non-natural possibilities took one outside the realm of science, for it violated the constraints of methodological naturalism. Methodological naturalism appears to be a commitment that science makes, namely to the ideas of testability and empirical evidence. It represents a basic value in science. Theories that require non-natural entities or forces, then, cannot be science.

Evolution by natural selection is virtually universally accepted by Western scientists and academics; support for literal interpretations of the Bible or similar accounts of the origins of humans is essentially nonexistent. On the other hand, a recent Gallup poll indicates that "46% of Americans believe in the creationist view that God created humans in their present form at one time within the last 10 000 years" (Newport 2012). The prevalence of this view has remained unchanged since Gallup started asking this question 30 years ago. It is rare that we find such a large gap between a well-received view in science and public opinion. And it is unclear how public educational institutions and the political infrastructure should deal with this gap, for it is virtually unique in contemporary society.

This debate is not between science and religion *per se*, as a committee formed by the NAS and the Institute of Medicine (IOM) charged with investigating this issue noted: "Today, many religious denominations accept that biological evolution has produced the diversity of living things over billions of years of Earth's history. Many have issued statements observing that evolution and the tenets of their faiths are compatible. Scientists and

theologians have written eloquently about their awe and wonder at the history of the universe and of life on this planet, explaining that they see no conflict between their faith in God and the evidence for evolution" (NRC 2008, 12). It is a debate between a few influential denominations who maintain a very conservative interpretation of the Bible and the scientific community (cf. Clark 2014). However, that the denominations are few has little bearing on the current depth of the conflict over how society should train future generations to think about naturalism and evolution in the contemporary world.

The methods of science belie any attempt to inject supernatural forces into theories and explanations. That said, and despite what the ontological naturalists maintain, there is nothing in science that would prevent one from ascribing supernatural forces to the universe. Evolution by natural selection is one example of this presumed tension. Some evolutionary biologists believe in God's Hand at work; others do not. From the perspective of science, both views are legitimate and compatible with the work that scientists do.

We get in trouble when either side of the naturalism/non-naturalism divide wishes to take the debate further and tries to use the methods of science (or the supposed failures of the methods of science) to prop up its favored ontology. Science cannot bear that burden, nor should it. For science is not an activity of certainty, but one of collecting data, making best guesses, and then, more often than not, being proven wrong in some fashion. Science gives us much, but it does not and cannot tell us whether God exists or, if He does, what His involvement in the unfolding of the world has been.

References

Balfour, A. (1906). *The Foundations of Belief: Being Notes Introductory to the Study of Theology*, 8th edn. New York: Longmans, Green.

Behe, M.J. (2001). Reply to My Critics: A Response to Reviews of Darwin's Black Box: The Biochemical Challenge to Evolution. *Biology and Philosophy* 16: 685–709.

Branch, G. (2006). Defending the Teaching of Evolution: Strategies and Tactics for Activists. In *Not in Our Classrooms: Why Intelligence Design Is Wrong for Our Schools*, edited by E.C. Scott and G. Branch, pp. 130–152. Boston: Beacon Press.

Branch, G. (2011). Introduction. *Synthese* 178: 171–176.

Brewster, H. (1862). The Facts and Fancies of Mr. Darwin. In *Good Words*, Vol. 3, edited by N. Macleod, D. Macleod, and H. Aspden. London: Alexander Strahan.

Clarke, K.J. (2014). *Religion and the Science of Origins: Historical and Contemporary Discussions*. New York: Palgrave Macmillan.

Darwin, C. (1860). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. (1872). *The Expressions of the Emotions in Man and Animals*. London: John Murray.

Darwin, C. (1881). Letter to William Graham. Available from: http://www.darwinproject.ac.uk/entry-13230 (last accessed July 15, 2015).

Darwin, F. (ed.) (1887). *The Life and Letters of Charles Darwin, Including an Autobiographical Chapter*. London: John Murray. Available from: http://darwin-online.org.uk/EditorialIntroductions/Freeman_LifeandLettersandAutobiography.html (last accessed July 15, 2015).

Dawkins, R. (1986). *The Blind Watchmaker*. New York: W.W. Norton.

Dawkins, R. (2003). *A Devil's Chaplain: Reflections on Hope, Lies, Science, and Love*. Boston: Houghton Mifflin.

Dembski, W.A. (1999). *Intelligent Design: The Bridge between Science and Theology*. Downers Grove: InterVarsity Press.

Dembski, W.A. (2002). *No Free Lunch: Why Specified Complexity Cannot Be Purchased without Intelligence*. Lanham: Rowman and Littlefield.

Dennett, D.C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.

Ecklund, E.H. and Scheitle, C. (2014). Religious Understandings of Science. Presentation at the American Association for the Advancement of Science, February 16, 2014, Chicago, IL.

Eldredge, N. and Gould, S.J. (1972). Punctuated Equilibria: An Alternative to Phyletic Gradualism. In *Models in Paleobiology*, edited by T.J.M. Schopf, pp. 82–115. San Francisco: Freeman Cooper.

Fetzer, J.H. (2007). *Render unto Darwin: Philosophical Aspects of the Christian Right's Crusade against Science*. Chicago: Open Court.

Fetzer, J.H. (2011). Evolution and Atheism: Has Griffin Reconciled Science with Religion? *Synthese* 178: 381–396.

Griffin, D.R. (2006). *Evolution without Tears: A Third Way beyond Neo-Darwinism and Intelligent Design*. Claremont: Process and Faith.

Gould, S.J. (1980). *The Panda's Thumb*. New York: W.W. Norton.

Herschel, J.E.W. (1861). *Physical Geography*. Edinburgh: Adam & Charles Black.

Kitzmiller v. Dover Area School District. 2005. 400 F. Supp. 2d 707 (M.D. Pa. 2005).

Levin, J.S. (1996). How Prayer Heals: A Theoretical Model. *Alternative Therapies in Health and Medicine* 2: 66–73.

Lewis, C.S. (1947). *Miracles: A Preliminary Study*. London and Glasgow: Collins/Fontana.

Linnaeus, C. (1762). *Fundamenta Fructificationis.*

Mayr, E. (1988). *Toward a New Philosophy of Biology*. Cambridge, MA: Harvard University Press.

Miller, J.D., Scott, E.C., and Okamoto, S. (2006). Public Acceptance of Evolution. *Science* 313: 765–766.

Newman, J.H. (1868/1971). John Henry Newman to J. Walker of Scarborough on Darwin's Theory of Evolution. In *The Letters and Diaries of John Henry Newman*, Vol. XXIV, edited by C.S. Dessain and T. Gornall, pp. 77–78. Oxford: Clarendon Press.

Newport, F. (2012). In U.S., 46% Hold Creationist View of Human Origins: Highly Religious Americans Most Likely to Believe in Creationism. *Gallup Politics*. Available from: http://www.gallup.com/poll/155003/Hold-Creationist-View-Human-Origins.aspx (last accessed July 15, 2015).

NAS (National Academy of Sciences). (1999). *Science and Creationism: A View from the National Academy of Sciences*, 2nd edn. Washington, DC: National Academy of Sciences.

NRC (National Research Council). (2008). *Science, Evolution, and Creationism*. Washington, DC: National Academies Press.

NSTA (National Science Teachers Association). (2000). The Nature of Science. Available from: http://www.nsta.org/about/positions/natureofscience.aspx (last accessed July 15, 2015).

NSTA (National Science Teachers Association). (2005). Survey Indicates Science Teachers Feel Pressure to Teach Nonscientific Alternatives to Evolution. Available from: http://www.nsta.org/about/pressroom.aspx?id=50377 (last accessed July 15, 2015).

O'Connor, R. (2003). The Design Inference: Old Wine in New Wineskins. In *God and Design: The Teleological Argument and Modern Science*, edited by N.A. Manson. London: Routledge.

Owen, R. (1860). Review of *Origin* and Other Works. *Edinburgh Review* 111: 487–532.

Pennock, R.T. (2005). *Kitzmiller v. Dover Area School District* Expert Report.

Pennock, R.T. (2011). Can't Philosophers Tell the Difference between Science and Religion?: Demarcation Revisited. *Synthese* 178: 177–206.

Plantinga, A. (1993). *Warrant: The Current Debate*. New York: Oxford University Press.

Plantinga, A. (2000). *Warranted Christian Belief*. New York: Oxford University Press.

Plantinga, A. and Tooley, M. (2008). *Knowledge of God*. New York: Blackwell Publishing.

Ray, J. (1692). *The Wisdom of God as Manifested in the Works of the Creation*. Available from: http://books.google.com/books?id=dxkHAAAAQAAJ&printsec=frontcover#v=onepage&q&f=false (last accessed July 15, 2015).

Ruse, M. (2004). The New Creationism: Its Philosophical Dimension. In *The Cultures of Creationism: Anti-Evolutionism in English-speaking Countries*, edited by S. Coleman and L. Carlin, pp. 175–192. Burlington: Ashgate Publishing Company.

Sarkar, S. (2007). *Doubting Darwin? Creationist Designs on Evolution*. Oxford: Blackwell.

Sarkar, S. (2011). The Science Question in Intelligent Design. *Synthese* 178: 291–305.

Shallit, J. and Elsberry, W.R. (2004). Playing Games with Probability: Dembski's Complex Specified Information. In *Why Intelligent Design Fails: A Scientific Critique of the New Creationism*, edited by M. Young and T. Edis, pp. 121–138. New Brunswick: Rutgers University Press.

Shanks, N. and Green, K. (2011). Intelligent Design in Theological Perspective. *Synthese* 178: 307–330.

Sober, E. (2004). The Design Argument. In *Blackwell Guide to the Philosophy of Religion*, edited by W. Mann, pp. 27–54. Oxford: Blackwell.

Temple, F. (1884). Of the Relations between Religion and Science. Available from: http://www.gutenberg.org/files/17194/17194-h/17194-h.htm#LECTURE_IV (last accessed July 15, 2015).

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainly: Heuristics and Biases. *Science*185: 1124–1131.

van Inwagen, P. (2003). The Compatibility of Darwinism and Design. In *God and Design: The Teleological Argument and Modern Science*, edited by N.A. Manson. London: Routledge.

Wilkins, J.S. (2009). *Species: A History of the Idea*. Los Angeles: University of California Press.

9

Naturalism and Antinaturalism in the Sociology of Science

DOROTHEA OLKOWSKI

Philosophers of science have often noted that naturalism in science arose out of "the struggle to free science or natural philosophy from its origin within a religious understanding of the world" (Rouse 2007, 62). The apparent similarity between religion and science arose from the medieval idea of God as a rational being who created a rational world. Humans, using scientific experiments, wished to understand the world as both godly and rational, and the aim of modern science came to be that of discovering the *unique* truth about the world, a truth that only God could know (Prigogine and Stengers 1984, 52). But by the late 19th century, empirical sciences had begun to assume a new perspective called "naturalism," which privileges empirically grounded research. This new perspective undermined not only the connection between science and religion, but also the role of philosophy in describing nature, science, and human psychology as, more and more, empirical research began to assert its primacy over philosophical theories (Rouse 2007, 63).

The move to naturalism is epitomized by the conflict between Robert Boyle (1627–1691) and Thomas Hobbes (1588–1679). Boyle is recognized as one of the first natural philosophers to clearly articulate and advocate the *experimental method*. He and his assistant, Robert Hooke, perfected the air pump so as to carry out experiments on the vacuum. Hobbes, a political philosopher, was also famous in his own time for work in optics, geometry, and natural philosophy. He believed in observation and reasoning, but not in experiments, and opposed the idea of a vacuum. Though Boyle's experiments were never completely successful, they were eventually successfully repeated by other experimenters, who used more advanced versions of the air pump. What was new in Boyle's approach was his use of laboratory experiments, which produce facts witnessed by reliable experts who are able to repeat them (Latour 1993, 18–22).

The response of some philosophers to this attempt to displace reasoning with scientific experiment was strong and clear. Gottleib Frege (1848–1925) is especially acknowledged for analytic epistemology, the conceptual analysis that explicitly rejects the experimental work of psychology and the concepts it produces (Kitcher 1992, 53–54). But the lack of success of philosophies associated with purely logical analysis, as well as objections to Edmund

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

Husserl's (1859–1938) pure phenomenology, which argued that only philosophy could "clarify and justify the meaning and validity of science," led eventually to the emergence of several varieties of what is now known as "philosophical naturalism" (Kitcher 1992, 56; Rouse 2007, 63). Though contemporary naturalism exists in a number of different forms, generally the point of naturalism is to replace philosophical speculation with empirical research; thus, to use science to carry out philosophical work (Rouse 2007, 63). This implies that philosophical accounts of science must now take their cue from actual scientific practices. Doing this leads, in turn, to the acknowledgement that philosophers must accept the scientific model of explanation, which is built on empirical causal relations and natural laws (Rouse 2007, 64).

Simultaneous with the rise of naturalism in the philosophy of science came an awareness and broader recognition of the influence of society on scientific inquiry, as well as what is referred to as "pluralism," the claim that natural entities and processes call for more than a single theoretical or investigative approach (Longino 2002, 1, 93). These approaches to philosophy of science have come to be known as *science studies* or, more commonly, *sociology of science*. Among the advocates of this approach, many consider themselves to be naturalists, but they tend to situate themselves at different points along a continuum, depending on how strongly they are committed to the ability or inability of philosophy to distinguish purely cognitive or theoretical approaches from social interests in science. Helen Longino defines naturalism in the social studies of science as "treating the conditions of knowledge production by human cognitive agents, empirical rather than transcendental subjects, as the starting point for any philosophical theory" (Longino 2002, 10). This means that the godlike view, the transcendental, is gone once and for all and empirical research takes its place.

The sociologists of science, such as Bruno Latour and Steve Woolgar, who represent one end of the continuum argue that it is impossible to differentiate the practices of science from the social and cultural influences that shape them. At the opposite end of this continuum, one finds philosophers such as Philip Kitcher, for whom naturalism means a minimal reintroduction of psychological factors into epistemology and a rejection of the concept of the *a priori*, the idea that there may be purely rational rules governing all experience. From his perspective, any attempt to engage in a serious reform of the language utilized to describe knowledge requires that philosophy of science first provide an account of the social dimension of human language (Kitcher 1992, 113). The work of Longino tends to fall in the middle between these two positions as she embraces aspects of both naturalism and antinaturalism. This essay will examine the tension between naturalism and the tenets of the sociology of science, which is predominantly antinaturalist, in order to determine to what extent, if any, the two are compatible.

Sociology of Science

The revolution in science studies or the sociology of science may be traced back to John Stuart Mill, Charles Sanders Pierce, and Karl Popper, each of whom pointed out, in different contexts, the fallibility and the social dimension of human knowledge and hypothesis testing. Nevertheless, it is more recent practitioners, who understand social construction much more broadly, who have received the most attention. These practitioners of the *strong program* in the sociology of science explain science with reference to ideological, political, professional, and individual career interests (Longino 2002, 11). Separately, there is a *weak*

program, whose advocates criticize science with respect to its human and environmental consequences, and who tend to be sociologists rather than philosophers. Advocates of the strong program – both philosophers and sociologists – argue that scientific theories, and even the natural world, are, in some sense, socially constructed. Thus, it is often referred to as an "ontological" view: the view that nature itself is sociologically constructed. Given the attention that this view has received, in spite of the fact that, in philosophy, it is a minority view, this essay will initially focus primarily on those theorists who understand social construction in this way, rather than epistemologically; that is, as characterizing the plausibility, worth, or truth of scientific work (Longino 2002, 12).

Laboratory Life

Perhaps the work of Latour and Woolgar gained prominence because their book, *Laboratory Life* (1979), an examination of laboratory activity at the Salk Institute in La Jolla, California, was published with an enthusiastic forward by Jonas Salk and supported with fellowships from Fullbright, NATO, and the Salk Institute, among others. In his introduction, Salk claims that Latour began with a sociological study of biology, but that his thought was transformed by biological concepts and ways of thinking, and, in fact, became a subset of the scientific activity, which is itself a subset of the process of life organization (Latour and Woolgar 1979, 13). The question is whether or not this statement holds true in Latour's methodology. Can it be considered a subset of biological concepts and the practices of biologists, which themselves are drawn from nature? If so, it can be defined as naturalism. Such an outcome would be ironic, given the major role played by Latour in the development of profoundly antinaturalist methods and conclusions in *Laboratory Life*. This makes the desire to examine Latour's work before moving on to other versions of science studies that are more comfortably affiliated with naturalism all the more compelling. The focus on this one account is justified insofar as it is the foundation of all of Latour's later work, and to a large extent becomes the model for antinaturalism in a growing number of contemporary philosophical projects.

The authors of *Laboratory Life* contend that much of what a sociologist may describe of scientific activity – even the scientific activity of the highly educated scientists of the Salk Institute – depends on routinely occurring minutiae. The sociologist's study of scientific activity is compared to that of "an intrepid explorer of the Ivory Coast" who has studied the *belief systems and material conditions* that produce the tribesmen's "savage minds" (Latour and Woolgar 1979, 28). Latour claims that he will collect and describe scientific activities in a particular setting, without an *a priori* hypothesis, and thus that the scientific activities will be approached as *strange*. This is partly so as to not distinguish between the technical and the social, and partly so as not to privilege the outputs of what he calls the "tribe" of scientists, nor take for granted technical culture (1979, 28–30). The work of the laboratory is described, following Thomas Kuhn, as "normal science": routinely dull work, lacking drama and conjectural daring, so as to focus on the "*social* construction of scientific knowledge… the *process* by which scientists make sense of their observations," insofar as even perception is said to be constituted by prevalent social forces (1979, 32, 33).

Careful attention is given to establishing which social–technical procedures allow the scientist to construct an ordered account out of what is called the "chaos" of possible perceptions, because "both scientists and observers are routinely confronted by a seething

mass of alternative interpretations" (1979, 36). The authors also claim that, in the research environment, any one interpretation can always be undermined by another, so every alternative may be questioned. This is why the elimination of alternative interpretations of scientific data is judged to be central to scientific activity. Order must eventually emerge out of the "seething mass" and "utter confusion" of disorder (1979, 36). Indeed, since one member of the institute, Roger Guillemin, shared a Nobel Prize for Medicine in 1977 for discoveries concerning the peptide hormone production of the brain, it seems that at least one key framework and some important interpretations were reached.

Latour claims to take up the point of view of an alien newcomer – not only to a laboratory, but to science all together – who encounters a "mysterious and apparently unconnected sequence of events" (1979, 38). The account purports to describe the experience of the anthropologist, but what about the work of scientists engaged in research in a naturalistic manner? If the stated goal of the study was to discern "the way in which the daily activities of working scientists lead to the construction of facts," then how the sociologist of science constructs an account of this process is relevant (1979, 40). The question is, do Latour and Woolgar succeed in producing an account that is an adequate description provided by an observer with an intermediary status, neither totally unfamiliar nor totally at home with scientific activity?

One way in which the scientists in the laboratory sort through the chaos is to make sense of "their juxtaposition of literatures by reference to a world of literature published outside the laboratory" (1979, 54). Latour and Woolgar refer to the ubiquitous presence and use of such scientific literature as "scripture" and to the practice of neuroendocrinology – the primary field of study at the Salk laboratory under observation – as having "all the attributes of a mythology," including mythical founders, like the ancient gods who founded Greece (1979, 54). Neuroendocrinology arose in the 1940s by bringing together neurology (the science of the nervous system) with endocrinology (the science of the hormonal system) (1979, 54). The story of how these two were combined when it was discovered that nerve cells secrete hormones and that there is no nerve connection between the central nervous system and the hormonal system is referred to as the "mythology" through which this scientific culture represents itself (1979, 54–55).

Unlike the scientists, who concern themselves with whether or not their work is correct, the anthropologists focus on the scientists' processes, including a set of "cultural values": the arguments and beliefs that form the basis of their work. For example, the hypothesis that the brain (especially the hypothalamus) controls the endocrine system by means of "releasing factors" is taken to be mythology because it is part of the so-called "culture," a set of arguments constantly appealed to that are noncontroversial among the scientists (1979, 55). The use of such concepts undermines the naturalist's claim that Latour's thought was transformed by biological concepts and naturalistic methods, and so became a subset of the scientific activity, which is itself a subset of natural processes. All the more so when Latour and Woolgar acknowledge that real-world breakthroughs, such as a process whereby a natural substance is modified and the modification is found to have "major implications for the treatment of diabetes," are nevertheless a "mythology" (1979, 61).

Latour and Woolgar defend their position by pointing out that the scientific literature, the product of laboratory work, quickly becomes the focus of discussion among scientists, and that the physical processes giving rise to it are forgotten or taken for granted (1979, 63). This, they argue, is the mythology that informs scientific activities. This is debatable insofar as the publishable results of experiment would not exist without the experiments producing

them, the so-called *material* setting of the laboratory (1979, 64). And yet, they think it is "wrong to contrast the material with the conceptual components of laboratory activity," because laboratory equipment is often the subject of debate in other published scientific literature, so the physical equipment of the laboratory is reducible to the concepts and theoretical arguments in another context (1979, 66). When a technician uses a computer, engaging with electronics to generate statistics, or a spectrometer is developed through spin theory, the implication is that the ideas and equipment should not or cannot be contrasted. Borrowing a term from Gaston Bachelard, every laboratory apparatus is a *reified theory*, physical and material equipment and processes transformed into concepts.

The transformation of laboratory processes into conceptual constructs is facilitated when Latour and Woolgar characterize scientific papers and statements as objects. If, for example, a peak on the spectrum of a chromatograph is observed more than once under independent circumstances, it is designated as a possible or likely "object." The scientists – according to Latour – are using the word "object" to refer to the peak on the chromatograph. But for Latour, "an 'object' was achieved through the superimposition of several statements or documents…outside of, or beyond, the reader's or author's subjectivity" (1979, 84). The term "author" does not refer to a person but to a *statement*; a single author is a statement coming from a single source, as opposed to one with several sources (1979, 84). Latour admits that, in a scientific paper focused on "results," the results have an external reference and independent presence, but, for him, this serves to produce no more than "an impression of objectivity" (1979, 85–86). In other words, the experimental work of the laboratory and the scientific papers are not separate. The sociological observers describe laboratory activity as the organization of persuasion through *literary inscription* whose outcome is the production of a fact – a fact that appears *to the scientists* to be beyond the scope of sociological and historical explanation, and so, for the scientists, cannot be claimed by the strong program (1979, 105).

Nevertheless, even if the scientists tend toward naturalism and insist on the distinction between written "inscriptions" and laboratory experiments with outcomes that exist independently of observers, Latour does not. For Latour, there is no difference between the processes by which substances are constructed (such as written documents) and those substances themselves. Moreover, when the scientists experimentally constitute a substance (e.g., a peptide referred to as TRF) by separating or extracting it from a relatively homogeneous material environment, for Latour they are not discovering a preexisting substance but creating something new (1979, 128–129).

This conflation between independent existence and preexistence, between creating something and discovering it, becomes problematic when a controversy arises over a hypothesis that turns out to be incorrect. Latour calls this a literary inscription, a material operation for creating order (1979, 245). When the inscription is eventually overturned by additional laboratory evidence, a new hypothesis is formulated. Latour claims that hypotheses are deductively logical and that the incorrect one must not have been derived from true premises. This overlooks the *evidence*, which is needed to disprove one hypothesis (e.g., that no substance had been constituted) and replace it with a different one (e.g., that a substance had actually been constituted). The new evidence requires scientists with different kinds of expertise from those who were initially working on the project, as well as additional equipment and instruments (e.g., a mass spectrometer, to provide information about the atomic structure of TRF). In the TRF case, this led to the conclusion that TRF is a peptide with a specific structure of amino acids (1979, 146–147). Thus, the determination

that TRF is a substance is in fact dependent on new evidence and is not a failure of logical deduction. Yet, Latour and Woolgar state that TRF is a "thoroughly social construction," and that deductive reasoning – applying general rules to particular cases – is no more than common sense, akin to facts, practical activities, the reliability of other researchers, and numerous related matters beyond the scope of experiments and experimental results (1979, 152, 158).

Because Latour emphasizes the purely social nature of discussions among scientists, it is difficult to distinguish what is social from what is an experiment or a hypothesis. Ideas – private thought processes – must be shared with the community and validated publically. The idea is no more than "a summary of a complicated material situation," and is always social (1979, 170). Likewise, the apparent correspondence between statements and independent objects taken to be *out there* actually "stems from the splitting and inversion of a statement within the laboratory context" (1979, 177). A statement is a set of words that represent a statement about an object that also corresponds to the object itself. Too often, the object is mistakenly assumed to be reality, while the statement is said to mirror that reality. All of this, says Latour, is an illusion. "When an accepted statement is doubted or qualified, the reality…melts back into a statement" (1979, 179).

The position is constructivist and antinaturalist. The argument is posed overtly in opposition to metaphysical realism, which can be described as "insisting that both the truth-or-falsity and scientific acceptance or rejection of any proposed law is independent of whatever factors may have conditioned the route to its proposal," so that discovery should not be confused with justification (Tiles 1991, 226). For Latour and Woolgar, this means that science is about activity and not about nature, that science simultaneously constructs reality and treats laboratory processes as reified theory (1979, 238, 242). It seems clear that this formulation excludes any version of naturalism as defined in the introduction.

Robust Social Epistemology

Unlike Latour and Woolgar, who remain staunchly antinaturalist, Philip Kitcher firmly situates the philosophical account of science as a naturalist and normative enterprise, so as to distinguish "those social processes and structures that enhance the acquisition of true beliefs or cognitive processes from those that do not" (Longino 2002, 71). In order to carry this out, Kitcher positions himself as a self-appointed "philosopher-monarch" who is organizing a scientific workforce for the promotion of the collective achievement of *significant truth*, even as he shifts attention from scientific *knowledge* to scientific *practices* (1993, 305). As Kitcher develops this idea, his model brings to mind nothing so much as Adam Smith's conception of labor, where "the greatest improvements in the productive powers of labour, and the greater part of the skill, dexterity, and judgment, with which it is anywhere directed, or applied, seem to have been the effects of the division of labour" (Smith 1776/1993, ch. 1).

For Kitcher, the production of significant truth seems to stand as a version of the concept of improvements in the productive powers of the workforce, though this workforce is making science, not pins, as in Smith's example. This analogy is encouraged by his statement regarding how to approach a scientific problem situation, which is to *optimize* the scientific community's response by means of the appropriate *distribution of the efforts* of individual members in order to ensure the success of that distribution in relation to the desired goal of equilibrium. "Equilibrium" refers to the avoidance of what he calls "cognitive disasters,"

similar, perhaps, to the avoidance of market disasters, such as when supply exceeds demand (Kitcher 1993, 304–305; Smith 1776/1993, ch. 7). To accomplish this, it seems initially to be necessary for the science workers to internalize the values of this particular monarch. They must be a "society of unrelated individuals, each of whom is pure of heart," since the alternatives are that they are either "motivated in baser ways" or "locked into systems of authority and deference," or both (Kitcher 1993, 305). Perhaps, then, the philosopher-monarch is operating just like Smith's invisible hand of the market, whereby unrelated individuals blindly pursuing their own self-interest end up serving the good of society as well, since society is governed by the economic law of supply and demand. Possibly this is why, on Kitcher's analysis, the distribution of labor in the science system ends up being organized in accordance with structures of authority and deference.

Kitcher proposes that authority is fundamental to the scientific endeavor. Science is a progressive activity, working its way free of past errors through "encounters with nature," while ensuring that, in the course of individual research, the mutual interactions of scientists do not impede "the process of self-correction" (Kitcher 1993, 306). Acknowledging that scientists in any lab will rely on others – technicians, graduate students, support staff – it follows that any opportunity that arises, from new information in a journal article to new equipment to a grant proposal, must be evaluated to determine whether the work should be carried out by the individual scientist or delegated to others (1993, 307–308). The question Kitcher raises is: How does the scientist decide how to proceed? How does the scientist decide whether to carry out the work herself or to obtain what she needs from other scientists who are also authorities?

As Kitcher points out, the success of a project relies on two probabilities: (1) the probability that the acquired information is correct and (2) the probability that it will lead to correct information that can be utilized at a second or later stage (1993, 308). In order to increase the likelihood that these probabilities are correct, Kitcher advises that, if the authority from whom the scientist is borrowing has greater expertise, the scientist should rely on that authority. However, under some circumstances, such as when the costs of borrowing are negligible, the scientist could tolerate a certain rate of error and proceed with the acquired information. The degree to which this is tolerable may be decreased by new information that rapidly produces results or increased if it requires the scientist to learn something, such as how to use this new information (1993, 309). What matters initially in this calculation is that the goal is to solve a particular problem and that the agent is what Kitcher calls *epistemically pure*.

This agent is contrasted with one who is *epistemically sullied*, a *scientific entrepreneur* whose primary motivation is to be the first to solve a scientific problem (1993, 310). Like Adam Smith's conception of laborers who are all fundamentally alike, this calculation assumes that no scientist has any greater talent than any others – or, if they do, that the error rate of the great talent is the average of those in competition with her. Perhaps surprisingly, the sullied scientists win over the pure ones because they are the ones who take advantage of "ample opportunities for division of cognitive labor," some borrowing from another authority and others working independently (1993, 311). Even though the term "sullied" is applied here, the scientists seem to be following the prescribed pattern for labor success; that is, the maximized division of labor. In addition, various asymmetries begin to emerge under these conditions. For example, some scientists are judged qualified to borrow others' work and some are not, and some scientists have few resources and so take bigger risks, while others have many resources and so can proceed conservatively.

As we can see, one important consideration underlying the determination that something is pure or sullied, a success or a failure, is the question of who is an authority. Authority can be of two types: unearned, meaning a result of social position, such as working at a major institution or having studied with a prominent scientist; or earned, meaning having a track record that others judge positively. Yet, the question remains: Does authority stem not only from prestige within the community, but also from personal relations between so-called authorities and those who evaluate their work (1993, 315)? Interestingly, it is in this context that the question of credit emerges. Credit is based on an overall assessment of a scientist's work, and Kitcher is very concerned with where this assessment arises from, as well as from whom it arises, since authority can be widespread, even mundane, and does not guarantee that those who possess it can grant credit.

Certainly, Kitcher expresses strong opposition to Latour's claim that authority arises from the social structure, so he maintains that scientists perform a kind of computation to determine who has authority, and he sets out this and other computations using equations. Yet, each of these computations can also be stated in ordinary language. The computation about authority is the equivalent of a comparison of one's own beliefs about what is true with the beliefs of others. In spite of the computations, it seems that the comparison of one's own beliefs with those of others is a social phenomenon, and the fact that it can also be expressed computationally is interesting but does not affect the fundamental principle at work here, which is the comparison of beliefs (1993, 316–317).

Similarly, comparison operates for the question of credit. If credit is given to scientists who obtain a *reputation* for knowing where the field is going, what is this based on? It is based on the judgments of one's peers, who compare one's views with their own. Kitcher claimed to admit at least some psychological factors into his analysis, but, again, this seems to be a social view. Prestige operates in the same way. A scientist with a "prestigious position" can easily block acceptance of a "new, controversial finding" from another scientist who lacks that prestige, even if other scientists agree on the evidence contributing to that new finding (1993, 318–319). Authority, credit, and prestige can also all be extended to authorities in other fields, whose expertise must be indirectly judged. Additionally, for Kitcher, there is no elimination of so-called "backscratching": the situation in which "your positive opinion of me raises my respect for you, which in turn may further increase your attribution of authority to me" (1993, 323). This, like the other determinations, seems to be an effect of social interactions, in which epistemic purity plays little or no part.

Given the role of authority, credit, and prestige, Kitcher examines the problem of being *perceived* by other scientists to be the first to be right, as opposed to being perceived to be wrong, and of being *perceived* to have been beaten by someone else who is perceived to be right (1993, 235). But the problem is complicated when the entrepreneurial scientist must decide whether or not to borrow work from another scientist without first determining whether it is accurate. The question is: What will other scientists think of this borrowing? Lacking extreme competition from other scientists, the entrepreneur has the leisure to check the results she is borrowing before proceeding. If there is no blame forthcoming if the solution is challenged, the entrepreneur may borrow or not borrow: the perception will be the same, unless of course the scientist who carried out the borrowed work is unpopular or *perceived* to be unreliable (1993, 327–328). In the end, not checking the borrowed work is risky, and the entrepreneur is likely to be blamed by other scientists in the community for not doing so. In situations where competition looms, we would expect – and, indeed, are informed that – whether or not the entrepreneur first checks the borrowed results depends on if she is "extremely self-confident"

but has a low status in the community, or on whether she has a low rank in the community in general. In the latter case, the scientist has little to lose except self-esteem.

Kitcher argues that the work of a scientific community governed by the philosopher-monarch seeking "the best chance of attaining the community ends of science" depends on the content of its findings and the trustworthiness of the source. Any attempt to replicate a finding calls for a very reliable authority if the cost of error is great (1993, 334–335). Nevertheless, Kitcher's philosopher-monarch errs on the side of conservation of well-established scientific claims, regardless of the reliability of the authority – again, unless there are replications by persons (more than one) of high authority and such replication costs the community little or nothing (1993, 336). But new findings that overturn current science always cost someone something, so it seems that science advances little or slowly under this model.

The Division of Cognitive Labor

In the first chapter of Kitcher's book, *The Advancement of Science,* he states that philosophical reflections on science stand in relation to the practice of science much as economic theory stands in relation to economic transactions involving work, money, and goods (1993, 10). Nevertheless, it is the task of philosophy of science to idealize or approximate phenomena and to recognize the rationality of individual scientists, as well as the rationality of the scientific community, without neglecting the details of scientific transactions. Thus, Kitcher develops equations that account for social moves, but those social moves can be understood and described independently of the equations. Their idealization requires computation, but Kitcher admits throughout that idealization is unlikely to be carried out. This brings into question the degree to which scientists actually engage in abstract calculations and seems to bring Kitcher much closer to Latour than he would like to be. Given this, it makes sense to conclude with a brief examination of the interaction of individual scientists and the community in order to see to what extent the work of individual scientists depends or does not depend on the social and sullied judgments of the community of scientists.

In the end, Kitcher shows great appreciation for what he calls *sullied communities* of scientists and provides an extended set of equations to explain that the division of labor in such communities depends on the size of the work force, the rate of progress of competing methods in proportion to the number of workers, and the anticipated or actual return on the investment of effort – though any actual return would not be possible to anticipate (1993, 346–347). He argues that *epistemically pure* communities that do not select their methods by examining how others proceed and possibly borrowing from them seem doomed to fail. By contrast, so-called "ruthless" rational egoists who maximize their chances of achieving personal goals and proceed according to these dictates are supposed to act on the probability that one method or another would increase their chances of success – something once again difficult, if not impossible, to predict until actual success arises.

Nevertheless, Kitcher suggests that the pursuit of fame and fortune might well allow a community to do far better than the epistemic purists unless the latter encounter truly optimal distributions and methods, as long as the community does not grow too large (1993, 350–352). This is referred to as "good news." Success requires an autocracy in which the lab chief controls the workers, hypothetically allowing for maximum flexibility, but also for unfettered utilization of other's work. Interestingly, Kitcher cites the "community" of James

Watson and Francis Crick as the primary example of this (1993, 352). He does not mention that they made use of the unpublished data and the DNA photography of Rosalind Franklin, without her knowledge or permission. Even though Kitcher argues that even members of a sullied community must remain devoted to "truth," the fact that mere acceptance of results weighs so heavily in achieving a reasonable distribution reveals that undertaking scientific work as a problem of labor distribution guided by the invisible but rational hand of the philosopher-monarch remains problematic.

Throughout his account, Kitcher focuses on rival theories, the primary paradigm with which he operates. Not only are there rival theories, there are also rival categories of scientists: the terminology he uses to differentiate high-status and low-status scientists is *aristos* versus *proles* (1993, 383). Even in consensus-building, the aristos are the decision-makers and the proles are excluded. One wonders if, from this perspective, Rosalind Franklin would not have been judged to be a prole. With the emphasis on authority, credit, and prestige, it is difficult to locate in Kitcher's account either metaphysical naturalism (reflecting a commitment to understanding mind, knowledge, and morality as part of scientifically comprehended nature, from an empirical basis) or scientific naturalism (the idea that philosophy follows exclusively from scientific work, permitting no impositions upon science that serve philosophical but not scientific ends).

Knowledge and Knowledge-Productive Practices

Given Kitcher and Latour's perspectives, it is Helen Longino's (2002) third way that stands out all the more powerfully as the means for resolving the naturalism/antinaturalism debate in the sociology of science. In place of Kitcher's binary of aristos versus proles and Latour's material apparatus, Longino posits the interdependence of knowing subjects. In place of Kitcher's sullied and pure epistemologies and Latour's constructivism, Longino posits contextualism: that rules and procedures are immanent to the inquiry. And in place of Kitcher's rival communities of scientists egoistically competing with one another and Latour's social determinism, Longino posits pluralism: the view that no single theory or model accounts for all real-world physical and biological processes. Without a third way, it seems likely that the ideal of naturalism will be lost to the authority of entrepreneurs or the tenets of the strong program, so that the explanation of scientific knowledge and practice will end up being identified with powerful social factors and ideological, political, professional, and individual career interests.

Longino points out that "the philosophical response to sociological approaches to scientific knowledge ranges from outrage to selective incorporation" (2002, 42). Those with the strongest objections to sociological approaches identify science narrowly with rationality, and identify the social with practices that exclude scientific rationality. As such, for Longino, the question is what does and does not constitute knowledge. Is knowledge identified solely with rationality, or are there other ways in which we may designate some thing to be knowledge? Longino differentiates three senses of the term "knowledge" in order to articulate how each is associated with a different point of view regarding empirical and social accounts. We can follow this reasoning in order to ascertain the role naturalism plays in each of these points of view.

According to Longino, knowledge in science can refer to "sets of knowledge productive practices," to "knowing, that is, the relation of a cognizer to some content," or to

"the outcome of knowledge-productive practices" (2002, 77). Each of these determinations is further complicated by its relationship to empirical procedures and philosophical concepts.

With respect to knowledge as sets of practices that produce knowledge, Longino differentiates what she calls the "empirically-minded sociologist" from the normative philosopher. Though both understand knowledge production to be the result of empirical research, the empirical sociologist takes the justified belief that something is known to be the result of accounts and representations accepted by a *community*, "distributing new accounts of natural processes in such a way as to effect general assent to those accounts" (2002, 79). Such procedures reflect the antinaturalist position taken up by Latour. By contrast, the normative philosopher recognizes and accepts only beliefs that arise from sense perception or inductive reasoning from evidence (usually both), generally inviting some version of naturalism.

Knowledge as knowing is a more complex relationship, one in which empirical sociologists like Latour have little interest, insofar as they tend to regard knowledge as a product of productive practices, as previously defined. Normative philosophers, however, apply rigorous standards to what can be called knowledge. First, a relation must be established between a subject or knower and some content. This is called "belief." Then, the content must stand in a certain relation to its object; ideally, the relation is that the content is true. But then, there must also be a relation between the subject who believes some content and whatever serves to justify that belief. Appropriately, this is called "justification." Justifications may vary from accepted foundations or grounds requiring no justification themselves to coherence with other beliefs, or they may be causally derived from sense data or experience (a traditional naturalist's position). Philosophers will haggle over how to define each of these terms, but each must play a role. The disagreement here between naturalists and antinaturalists is about what constitutes legitimate belief: whether it is the subject's relations to a community or a set of relations independent of any community (2002, 80–82).

The relevant distinctions when discussing knowledge as content have to do with fact versus fiction, the correctness of one historical representation over another (such as relativity over classical physics), and the question of whether truth is valued over belief. What matters most in this instance is the third distinction: truth versus belief. For the sociologist, this once again concerns the role of community practices, but for the normative or conceptual philosopher, truth falls back on to the relation of the truth of some content to an object of investigation (2002, 83). The distinction between these two perspectives, Longino argues, is mostly a matter of whether or not one emphasizes the term "knowledge" or the term "production," as well as the conditions that define a subject's knowing or their membership in a community; finally, there is the historical, social, and geographical circumscription of some content or the performance of a logically sound argument (2002, 85–87). Though these differences of emphasis may well be operating, it also seems that the antinaturalism of the sociologists largely accounts for the difference in their methods, just as it is the naturalism of the normative philosophers that guides their thinking about knowledge.

Beyond Binaries

For her part, Longino argues for a third position, an alternative between the social–rational dichotomy and its underlying binaries, a third way that would dissolve this conflict between opposite positions. Perhaps it is not surprising, then, that Longino situates her work as both

antinaturalist *and* naturalist. It is not naturalist, insofar as she does not take knowledge to be a natural kind, something completely independent of humans, something whose nature can be discovered only by utilizing scientific or philosophical methods. But it is naturalist, insofar as it takes the starting point for knowledge to be the conditions of knowledge production by empirical, human, cognitive agents, and it takes knowledge to be a normative concept (Longino 2002, 10). Arguing for a third way, between binary oppositions, in place of the binary of community assent *or* individual cognitive agents, Longino posits the interdependence of knowing subjects. In place of the binary of constructivism – that natural processes are produced by empirical, scientific activities and need not correspond to anything real, and the monist position that there is only one correct, complete, consistent account of nature – Longino posits pluralism. This is the position that, since no single theory or model accounts for all the physical and biological processes of the real-world system, multiple accounts are necessary. And finally, in place of the binary of justification as arbitrary, or as impossible (relativism), if not unnecessary, because it is intuitive, and the claim that justification is context-independent, Longino posits contextualism: that justification depends on rules and procedures immanent to the context of inquiry (2002, 90–92).

Ultimately, it seems that pluralism, the third position between constructivism and naturalism, requires what Longino calls a "modest epistemology": one that does not promise complete knowledge and certainty insofar as it allows for social distinctions *as well as* normative judgments within its discourse (2002, 95). A modest epistemology recognizes that though naturalists argue that science is a function of how the world is and antinaturalist constructivists seem to say that science depends on human intellectual equipment and interests, in the end, epistemology cannot be used to completely rule out one or the other (2002, 95).

References

Dolphijn, R. and van de Tuin, I. (2012). *New Materialism: Interviews and Cartographies*. Ann Arbor: Open Humanities Press.

Kitcher, P. (1992). The Naturalists Return. *Philosophical Review* 101(1): 53–114.

Kitcher, P. (1993). The Advancement of Science: Science without Legend, Objectivity without Illusions. Oxford: Oxford University Press.

Latour, B. (1993). *We Have Never Been Modern*, translated by Catherine Porter. Cambridge, MA: Harvard University Press.

Latour, B. and Woolgar, S. (1979). Laboratory Life: The Social Construction of Scientific Facts. Beverly Hills: Sage.

Longino, H. (2002). *The Fate of Knowledge*. Princeton: Princeton University Press.

Prigogine, I. and Stengers, I. (1984). *Order Out of Chaos: Man's New Dialogue with Nature*. New York: Bantam.

Rouse, J. (2007). Naturalism and Scientific Practices: A Concluding Scientific Postscript. In *Naturalized Epistemology and Philosophy of Science*, edited by C.M. Mi and R. Chen, pp. 61–86. Amsterdam: Rodopi.

Smith, A. (1776/1993). *The Wealth of Nations*. Oxford: Oxford University Press.

Tiles, M. (1991). *Mathematics and the Image of Reason*. New York: Routledge.

10

Why Methodological Naturalism?

HANS HALVORSON

We are the beneficiaries of the efforts of many generations to unlock the mysteries of nature. The best scientific theories of today have achieved unparalleled predictive success, and this success provides some evidence that these theories have latched on to the structure of reality.

Why has science been so successful? What are some defining characteristics of the scientific approach, and of its products: scientific theories? One striking feature of these theories is that they are *naturalistic*: they don't mention gods, demons, or any other supernatural beings. Does the absence of supernatural beings from our best scientific theories provide evidence for the absence of such beings from reality? Contrapositively, should a person who believes in supernatural entities, most notably a theist, find scientific theories to be inadequate if they don't include these entities?

I will argue that the answer to both questions is "No." In particular, I will argue that there are good reasons – most especially for theists – to restrict scientific investigations to the natural world. Consequently, the success of methodologically naturalistic science does not in any way detract from the plausibility of supernaturalist theism.

The structure of this chapter is as follows. I will begin by undertaking a preliminary discussion of the aims of science. I will then consider how to define methodological naturalism, before discussing two motivations for adopting this approach. Finally, I will defend methodological naturalism against the criticism that it would impede the progress of science.

The Aims of Science

Many discussions about naturalism in science are marred by false presuppositions about the aims of science. These discussions often seem to assume that science aims at nothing less than discovering all truths. But it's immediately apparent that this cannot possibly be

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

the aim of any individual scientist, nor of the scientific enterprise as a whole. For example, no scientist is aiming to discover truths about my ice cream preferences, or about what I ate for breakfast this morning, or about what is morally right and wrong.

So if science isn't trying to discover all truths, then what is it trying to do? My hypothesis is that science aims to discover *certain sorts* of truths – in particular, those truths that can be systematized according to general schemata.

AIM: Science aims to explain a wide range of phenomena by unifying them under general schemata.

Let me make a few clarifications. First, I am thinking of AIM as a necessary, but not sufficient, criterion of scientific inquiry. Second, I have no opinion about the extension of the word "science." In fact, I would be happy enough to restrict my discussion to the scientific field that I know best: mathematical physics. If there is some science that doesn't satisfy AIM, then I will happily grant that my argument doesn't apply to that science. Finally, I'm using the term "general schemata" so as to be ecumenical between more specific views about the aims of science. The two paradigm cases of general schemata, however, are (1) statements of natural law and (2) mathematical models.

In the next section, I will argue that if AIM is true, then a theist should be a methodological naturalist. In the following section, I will argue that if AIM is true, then methodological naturalism doesn't impede the progress of science: it merely prescribes a good strategy for achieving the aims of science. In the remainder of this section, I'll provide some support for AIM.

Why should we believe that AIM is true? That is, why should we believe that science aims to explain diverse phenomena by unifying them under general schemata? First, I can speak from personal experience: when doing mathematical physics, we don't try to discover every truth about the systems we're studying. Rather, we try to construct mathematical models that capture some of the most salient structural features of these systems.

Second, and more importantly, AIM is a natural consequence of the views of some of the most acute philosophers of science. First, Immanuel Kant's view of science certainly validates AIM: "The single most distinctive criterion of demarcation for science, according to Kant, is *systematicity*" (Watkins 2012, xiv). Second, Hempel and Oppenheim (1948) argue that science aims to discover natural laws and then to use these natural laws to explain individual events. Again, what is characteristic of scientific knowledge is its bringing individual facts together under general schemata. Third, van Fraassen (1991) contends that science aims at constructing empirically adequate models – in other words, at fitting the phenomena into a mathematical scheme that facilitates prediction and understanding. Finally, in a recent book, Hoyningen-Huene (2013) argues that systematicity is in fact the feature that distinguishes scientific knowledge from other forms of knowledge. This list of philosophers of science could go on at length. But these examples suffice to show that some of the most insightful interpreters of scientific practice have claimed that science is interested in a particular sort of knowledge, namely knowledge of principles, laws, structure, or something of that nature.

So, experts recognize that science aims to systematize facts. But we didn't even need to ask experts, because we all know that science is in the business of constructing theories. And constructing a scientific theory is clearly distinct from describing reality. Of course, these two activities sometimes run together, since one of the main reasons for constructing

scientific theories is to use them to describe reality. But the two activities are not identical, since it's possible to describe reality without constructing a scientific theory. For example, if I say that "there is a lamp sitting on my desk," I am describing reality, but I am not constructing a scientific theory. Even if a scientific theory were a collection of sentences, it wouldn't be just any collection of sentences. Rather, it would have to be a collection of sentences that had a certain tight order and structure.

But if our goal is to describe reality, then why take this detour through constructing scientific theories? Why not just describe reality directly? Well, it seems that the point of theory construction is that it gives us a *systematic* way of arriving at truths about reality. Consider an analogy: If you were a prehistoric human, and your goal was to get some food, then why would you take the detour through planting crops? Why not just go out and forage for food? Of course, the reason for planting crops is so that you'll have a systematic method for producing food. Similarly, the goal of science isn't just to discover truth; rather, insofar as it relates to discovering the truth, it is to construct a system of truths.

Consider an example. Despite what you might hear in popular presentations, Einstein's general theory of relativity isn't simply a list of claims, such as "energy and mass are interconvertible." Rather, general relativity provides a collection of models that can be used to describe situations in which gravity is the predominant force. So, for example, general relativity provides a model of the overall expansion of the universe after the Big Bang; it provides models for collapsing stars; and it provides the models that we use to describe the orbits of GPS satellites around the earth. General relativity shows that these situations have common structural elements; it systematizes these diverse phenomena, providing us with an efficient means of generating predictions.

Defining Methodological Naturalism

A number of philosophers and scientists argue that science presupposes methodological naturalism – but that adopting methodological naturalism doesn't require commitment to metaphysical naturalism (see de Vries 1986; Scott 1993; 2004; Pennock 1999; 2011; Haught 2004; Ruse 2005; Miller 2009; Sober 2010; 2011). In this section, I will consider how to define methodological naturalism. In the section that follows, I will consider motivations for adopting it.

A typical strategy for defining methodological naturalism might proceed as follows: (1) define "*x* is natural"; (2) define *metaphysical naturalism* as the belief that "everything is natural"; (3) define *methodological naturalism* as the strategy of acting *as if* metaphysical naturalism were true. But this way of thinking about methodological naturalism has at least two major flaws. First, there are problems with defining "*x* is natural." Second, there are problems with thinking of methodological naturalism along the lines of acting *as if* metaphysical naturalism were true.

Natural Things

Methodological naturalism requires scientific theories to mention only natural things. One problem with this suggestion is that scientists are constantly postulating new entities, such as quantum wavefunctions, quarks, and genes. And who is to say whether or not these entities

are natural? What are the defining characteristics of natural entities? The problem before us, then, is to complete the following definition:

NAT: *x* is *natural* just in case…

But it would be extremely difficult to complete this definition in a way that would be useful for guiding scientific practice.

First, it wouldn't be helpful to define natural entities as those mentioned by our *current* best scientific theories. This is because methodological naturalism would then lead to extreme conservatism about ontology: no new entities should be introduced in science.

Second, it wouldn't be helpful to define "*x* is natural" in terms of the words "natural," "supernatural," or any synonym thereof. For example, it wouldn't be helpful to define natural entities as those that are governed by natural laws, or as those that are studied by the natural sciences. It also wouldn't be helpful to define natural entities as those that are *not* supernatural, or as those entities that do *not* transcend nature.

Third, it wouldn't be helpful to define "*x* is natural" in terms of space, time, energy, or mass. Contemporary science already defies simple intuitions about what is natural, and we can expect future science to do so to an even greater extent. For example, quantum wavefunctions don't live in space and time, and yet most realist interpretations of quantum theory treat the wavefunction as a genuine natural entity. Similarly, photons are natural entities that neither have a location in space nor have mass. What's more, in trying to unify quantum mechanics with the theory of relativity, physicists are currently entertaining theories in which space and time themselves emerge from a more fundamental substratum of nonspatiotemporal, and yet fully natural, entities.

I'm not the first person to point out these problems in defining "natural entity." Rea (2002) and van Fraassen (2002), among several others, argue that metaphysical naturalism has no precise definition and so cannot be thought of as a hypothesis for which we could gather empirical evidence. At best, metaphysical naturalism is an attitude, stance, or research program. I agree with the critiques of these philosophers. However, I'm not convinced that there is no useful heuristic distinction between natural and supernatural entities – a distinction that could be useful in helping delimit the domain of scientific inquiry. In fact, it seems that theism has a rough and ready answer: an entity *x* is natural just in case *x* was created by God.[1]

Provisional Atheism

Suppose, contrary to the argument of the previous subsection, that we had a clear distinction between natural and non-natural entities. In this case, a metaphysical naturalist would be a person who believes that everything is natural. And one might suggest that methodological naturalism requires us to do science *as if* metaphysical naturalism were true.

But this proposal doesn't seem to capture the intuition behind methodological naturalism. First, it confuses focusing attention on one sort of object with pretending that other

1 Theists might worry that this definition would classify angels as natural entities. My response: So what? As far as I can tell, theism doesn't need to classify angels as supernatural.

sorts of objects don't exist. To see that this is indeed a confusion, consider that theistic mathematicians don't mention God in their proofs, but presumably not because they are provisionally atheistic. Rather, these mathematicians are simply focusing on mathematical objects and the relationships between them.

Similarly, a typical methodological strategy in natural science is to delimit the domain of objects of study, or to delimit those aspects of things that will be studied. For example, in gravitational physics, attention is focused on those aspects of objects that can be represented in terms of geometric structures. But focusing on these aspects hardly amounts to denying that the objects of study might have other, nongeometric aspects, or that there might be other objects that don't admit of any geometric representation.

There is a yet more fundamental reason for thinking that methodologically naturalistic science isn't tantamount to provisional atheism: a theistic viewpoint played a crucial historical role in the development of the characteristic strategies of modern science, including the strategy of restricting focus to the natural world. The connection between theism and the development of modern science has been elaborated by Reijer Hooykaas (1972; 1987), Eugene Klaaren (1977), and John Hedley Brooke, among several others (see Clark 2014). The connection between theism and methodological naturalism is more than just a historical accident. The idea that a rational being (viz. God) created the universe by means of a free act suggests both that the universe has intelligible structure and that this structure can only be discovered by means of empirical investigation (rather than by *a priori* deduction from the concept of God). (An interesting presentation of the conceptual connection between theism and scientific method can be found in a series of three papers by Michael Foster (1934; 1935; 1936).)

Finally, there is some reason to worry that atheism actually runs contrary to the spirit that animates (methodologically naturalistic) natural science. Most poignantly, if the universe is ultimately purposeless, then why would one want to spend their short life working hard to understand it? Or following Plantinga's (2009) evolutionary argument against naturalism, if you assume that your cognitive faculties are the result of unguided evolutionary processes, then don't you have reason to be suspicious of the results of the scientific enterprise?

The previous paragraph should be taken with a grain of salt, because I don't intend to argue that if God doesn't exist, then there's no reason to do science. I only claim that provisional atheism, unlike methodological naturalism, isn't particularly well suited to helping science achieve its aims. In the next section, I will turn this point on its head by showing that theism provides a good motivation for methodological naturalism.

Motivating Methodological Naturalism

The previous section ended in *aporia*: we didn't find an adequate definition of methodological naturalism. So let's take a different tack. Let's postpone trying to define methodological naturalism until we better understand its motivation. In this section, I will discuss two possible motives for methodological naturalism. First I will discuss a proposal by Boudry, Blancke, and Braeckman (2010) to the effect that methodological naturalism is motivated by the history of failed supernatural explanations. Then I will discuss the original theistic motivation for methodological naturalism.

Intrinsic versus Provisional Methodological Naturalism

Boudry, Blancke, and Braeckman (2010) distinguish two versions of methodological naturalism, which they call *intrinsic MN* and *provisional MN*. The distinction turns on whether methodological naturalism is thought to be a presupposition of science (intrinsic) or simply a strategy that has proven to have good results (provisional). Boudry et al. object vehemently to intrinsic MN, claiming that it undermines the cause of naturalism: it lays defenders of methodological naturalism open to the accusation that they dogmatically exclude the supernatural from science. In contrast, provisional MN is supposed to be warranted in the same way that other scientific hypotheses are warranted: viz., by induction on past cases. Indeed, Boudry et al. claim that science provides strong evidence that there are no supernatural entities.

There are many problems with Boudry et al.'s argument for provisional MN. First and foremost is the fact (previously discussed) that the concept "natural" has a shifting meaning – thereby trivializing inductive arguments for the claim that "successful scientific theories will include only natural entities." Consider an analogy. Suppose that I have a hypothesis: All professors of philosophy are intelligent. But when I learn that Professor X took an IQ test and scored 90, I conclude that IQ tests must not be the sole criterion of intelligence. That is, rather than abandon my original hypothesis, I instead enlarge the meaning of "intelligent." In a similar fashion, at any time in history, it might be correct to say that our best scientific theories include only natural entities. But when the next scientific revolution comes along, we will enlarge the concept of "natural entities" to include whatever is described by the new successful scientific theories. That is, we say that something is natural as soon as it appears in a successful scientific theory.

Conversely, consider the claim that supernatural explanations have typically not been successful. There are two ways to understand that claim: one grandiose way, which is implausible, and one more modest way, which doesn't support provisional MN. According to the grandiose version of the "thesis of failed supernatural explanations," primitive peoples attempted to explain puzzling phenomena in terms of supernatural agents. But time and again, natural explanations (of the same phenomena) have proven superior, and they have therefore replaced those failed supernatural explanations.

This sweeping claim – with its echoes of Comtean positivism – oversimplifies the issues in many ways. First, it's not clear that there is all that much overlap between the events that science explains and the events that might be thought to call for a divine explanation. For example, science might explain why a certain liquid freezes at a certain temperature, but most religious creeds don't offer an account of phase transitions. Similarly, many religious creeds offer an explanation of why humans are valuable, whereas science seems only concerned with why humans might think that they are valuable.

Even in cases where science and religion purport to explain the same phenomena, people often have different intentions when offering a theological explanation than when offering a scientific explanation. For example, suppose we want to explain why Saul stopped persecuting Christians. Well, what are we trying to accomplish with our explanation? Are we trying to give a general account of how human dispositions are related to underlying physical conditions? Are we trying to describe Saul *qua* physical object? In that case, we might want some sort of neurological explanation. But perhaps we want to understand Saul's change in a way that we can relate to as persons; that is, in a way that makes sense from a motivational point of view. In that case, we might want an explanation that involves

agents – perhaps even agents who can (as some theists believe of God) stimulate transformations of a person's dispositions.

There is, however, a more modest version of the "thesis of failed supernatural explanation" that I will grant: theistic accounts of events often lack those features that we have come to prize in our best scientific accounts of events. Our best scientific accounts manage to systematize huge swathes of data under a few general schemata. These accounts are precise; in the best case scenario, they are couched in fully rigorous mathematical language. But must every good explanation have these features? "Sally went to the store because she wanted some bread." "John didn't accept the gift because it would have been wrong." These explanations lack several of the hallmarks of our best scientific explanations, and yet, depending on the context, they might be perfectly acceptable. Thus, theistic explanations may not be good as scientific explanations, but they needn't be intended as *scientific* explanations.

It seems, then, that the argument for provisional MN makes a couple of mistakes: first, about the concept "natural," and second, about what we have learned from the history of science. If there's something right about MN, it will be more along the lines of intrinsic MN; that is, as a presupposition of science. Is there any independent motivation for the strategy of restricting focus to natural things?

The Theistic Rationale for Methodological Naturalism

If you look more closely at the history of science, you will find that methodological naturalism wasn't born out of a metaphysically naturalistic ideology. Quite to the contrary, methodological naturalism arose in an explicitly theistic context, as an outworking of the doctrine of creation.[2] According to the Judeo-Christian doctrine of creation, our universe is an artifact – both designed and brought into being by God – and therefore it was built according to a blueprint that can be discerned by rational creatures like ourselves. Moreover, since God's choice of a universe was free and unconstrained by any natural law, the only way to discover the blueprint of creation is by means of empirical investigation. Therefore, the early modern scientists – all of whom were theists – believed that the following would be a worthy pursuit: to use one's intellect to put forward a possible blueprint of the natural (i.e., created) world; and then to use one's senses to test whether the proposed blueprint matched the actual universe.

In this chapter, when I speak of theism, I mean to include the claims: (1) the universe exists because God freely chose to create it; (2) if there are laws of nature, then God decided what they would be, and He is not bound to obey them; and (3) God interacts with creation, and is directly responsible for certain particular events in history.

If a person is a theist in this sense, then shouldn't they think that God and God's activities are a perfectly legitimate subject for scientific inquiry? I don't think so. In this section, I argue that theists should see natural science as aiming to reconstruct a blueprint of the universe. Just as the blueprint for a building omits reference to the building's architect, so the blueprint of the universe can be expected to omit reference to God.

The key idea behind my argument is the claim that I called "AIM": science aims to systematize phenomena by placing them under general schemata. If science's aim were to

2 For more on the origin of methodological naturalism in the work of Christian scientists, see Bishop (2013).

discover all truths, then a theist would certainly think that science ought to say something about God. But science doesn't aim to discover all truths; it aims to organize facts under general schemata (AIM). A theist might reasonably think that facts about God aren't of this sort, and so properly fall outside the domain of scientific systematization. I will argue for this claim by considering the two paradigm cases of general schemata: (1) statements of natural law and (2) mathematical models.

God is not Subject to Natural Laws

According to the classic view of Carl Hempel, science aims to discover natural laws and then to explain phenomena in terms of these laws. Now, Hempel's view can be challenged on various grounds. But my goal in this essay is not to defend a specific philosophy of science; it is to argue that in many scientific contexts, a theist has good reason to pursue a methodologically naturalist strategy. And the claim now is: to the extent that science is trying to explain things in terms of natural laws, a theist should be a methodological naturalist.

According to theism, the laws of nature are not metaphysically necessary; indeed, God could have chosen different laws of nature if He had wished. But this means that God's actions – unlike the actions of any created thing – aren't necessarily governed by the laws of nature. I will argue that since God isn't subject to the laws of nature, "God" shouldn't be a theoretical term in a scientific theory.

In a scientific theory, theoretical terms – such as "mass" or "Higgs boson" – are connected to each other by means of law statements. If a term *X* is not connected to other terms by means of natural law statements, then *X* is not properly part of that theory. But theists believe that "God" is not connected to other terms by means of natural law statements. Therefore, theists have good reason to suppose that God won't be mentioned in a good scientific theory.

Incidentally, these considerations suggest a reason why theists might be uncomfortable with intelligent design theory. According to intelligent design, there are certain natural phenomena – so-called irreducibly complex phenomena – that can only be explained by the action of intelligent agents. But it seems then that intelligent design must be committed to a statement of the form:

For any *x*, if *x* is irreducibly complex, then there must be an intelligent agent *y* that causes *x*.

What is the force of the "must" in this statement? If the statement is a law of nature, then the variable "*y*" ranges over *natural* things. But then the observation of an irreducibly complex phenomenon would indicate the existence of a *created* intelligent agent, not a creator.

Misrepresenting God

Even if science isn't searching for laws of nature, there are still reasons to think that "God" shouldn't appear as a theoretical term in a scientific theory. I will now consider a view that has been prominent in philosophy of science since the 1970s: the so-called "model-theoretic" view of science.[3] According to this view, science aims at constructing mathematical models that represent its objects of study. I will argue now that, given the model-theoretic view of science, a theist has reason to be a methodological naturalist.

3 A version of this view has been advocated by Ron Giere, James Ladyman, Elizabeth Lloyd, Fred Suppe, Patrick Suppes, Paul Thompson, and Bas van Fraassen, among many others.

Assuming that the model-theoretic view of science is accurate, a theist is faced with a question: Should science aim at constructing models that include mathematical objects (such as a number, or a vector, or a tensor) representing God, and that describe God's actions in the same manner that they describe physical processes (e.g., by means of differential equations)? Now, before you decide how a theist should answer this question, recall that the word "model" has several meanings, some quite general, and some quite technical. On the one hand, we can speak loosely of "a model of X" in any case where we construct some sort of representation of X. In this loose sense, of course, many theists will think it a reasonable aim to construct models of God, and of God's relationship to the universe. On the other hand, the model-theoretic view of science uses the word "model" in a technical sense, as a certain sort of mathematical structure. For example, in Einstein's theory of relativity, a model of the universe is a four-dimensional Lorentzian manifold. But if we take "model" in this technical sense, then theists shouldn't want to construct models of God, or of God's interactions with the universe.

Why shouldn't they? If God exists and interacts with the universe, then why shouldn't we be able to represent God and His activities via mathematical structures – just as we represent natural entities and processes? This is a good question. But perhaps a better question is: Why do we represent natural entities and processes by means of mathematical structures? The simplest answer to this is: It works. But for a theist, a more profound sort of answer suggests itself: We can represent the structure of physical reality using mathematics because the universe is an *artifact*, the creation of an intelligent mind. In particular, God created the universe according to a mental blueprint; and it is this mental blueprint that scientists aim to represent by means of mathematical models. So, for a theist, the applicability of a mathematical structure to physical reality depends on the fact that physical reality was created by a mind.

But now to return to the first question: Why shouldn't the goal of science be to construct a mathematical model of everything that exists – including God, if God does exist? Why shouldn't we aim to model God and God's activities? Well, why should we? After all, God isn't like the physical universe: He wasn't created; He wasn't first conceived in a mind and then brought into being. So, for a theist, the metaphysical status of the physical universe – viz., that it's an artifact – suggests that it has a blueprint that can be represented by means of mathematical structures. But a theist believes that God is a free agent, governed by no higher laws, and not Himself an artifact. God wasn't created according to a blueprint, and so there is no reason to think that God could be modeled by means of mathematical structures.

There are advantages and limitations to the method of representing things by means of mathematical structures. When we represent an entity X by means of mathematical structures, we provide ourselves with very detailed quantitative information about X. But the tradeoff is that some concepts don't yield to mathematical representation. For example, a geometrical description of my wife's hands provides me with detailed information about the size of her fingers. But a geometrical representation of her hands disregards their color and softness, not to speak of more high-level aesthetic qualities. In a similar fashion, if we attempted to represent God by means of a mathematical structure, then we would be committing ourselves to claims about God to which we are not entitled; and, what's worse, we would be shifting focus away from the information that theists think God has revealed about Himself (e.g., that He is just and loving). In other words, to represent God as a mathematical object would be to replace the revealed concept of God with a concept of our own making.

According to the Judeo-Christian-Islamic tradition, God revealed himself through the words of ordinary people – not through abstract, mathematized languages such as those developed for scientific purposes. Thus, these traditions' claims about God and God's relation to the universe are wrapped up in natural language, with all its ambiguities and emotive involvements. In contrast, the language of science has been tailored to maximize empirical informativeness, to eliminate ambiguity, and to minimize emotive involvement with the subject matter. As such, the language of science isn't particularly well suited to talking about other persons, including divine persons, if they exist.

Defending Methodological Naturalism

Methodological naturalism has detractors on both ends of the ideological spectrum: theists who argue that God shouldn't be kept out of our scientific theories, and nontheists who argue that the nonappearance of God in our best scientific theories is good evidence that God doesn't exist. In both cases, critics claim that methodological naturalism would disempower science – in particular, by denying it the power to address ultimate questions, such as "Does God exist?" and "Did God create the universe?" In this section, I will explain what's wrong with these criticisms of methodological naturalism.[4]

According to William Dembski:

> Although methodological naturalism is a regulative principle that purports to keep science on the straight and narrow by limiting science to natural causes, in fact it is a straightjacket that actively impedes the progress of science. (Dembski 2004, 170)

Similarly, Brad Monton claims that:

> If science really is permanently committed to methodological naturalism, it follows that the aim of science is not generating true theories. Instead, the aim of science would be something like: generating the best theories that can be formulated subject to the restriction that the theories are naturalistic…I maintain that science is better off without being shackled by methodological naturalism. (Monton 2009, 58)

Evan Fales opines that science shouldn't be limited to natural causes:

> A fundamental mission of science is to discover the causes (more generally, the explanations) of things. If there are supernatural causes, then science should seek them. (Fales 2009)

Finally, Fishman and Boudry claim that methodological naturalism imposes artificial constraints on science:

> Science, at least ideally, is committed to the pursuit of truth about the nature of reality, whatever it may be, and hence cannot exclude the existence of the supernatural *a priori*…without artificially limiting its scope and power. (Fishman and Boudry 2013, 921)

4 For some more nuanced criticisms of methodological naturalism, see Plantinga (1997), Ratzsch (2004), and Koperski (2008). I think these are effective against certain versions of methodological naturalism, but not the one for which I've argued.

> …MN imposes artificial constraints on science which are antithetical to its fundamental goal: to pursue the truth about the nature of reality on the basis of the evidence, wherever it may lead. (Fishman and Boudry 2013, 923)

These are some formidable opponents for methodological naturalism. What are we to make of their criticisms?

These criticisms have a common structure, which we can represent as follows:

(1) Science aims to *X*.
(2) Requiring methodological naturalism might hinder science from *X*-ing.
(3) Therefore, methodological naturalism should not be required.

The *X* varies slightly from critic to critic. For example, Monton indicates that science aims to generate true theories, while Fales suggests that science aims to discover the causes of things. I will argue, however, that there is no *X* such that science plausibly aims at *X* while methodological naturalism hinders science from *X*-ing. In other words, I will argue that premises (1) and (2) cannot be simultaneously true.

Let's begin with Monton. In this case, I'll grant the first premise: that science aims to generate true theories. But what is a scientific theory? As Monton well knows, a true theory is not just a collection of facts – and so, saying that science aims at generating true theories is not the same as saying that science aims to generate facts. For example, if the semantic view of theories is true, then a theory is a collection of (mathematical) models, and a theory is true just in case one of its models is isomorphic to its intended domain. In this case, science aims to generate models that are isomorphic to the intended domain.

Could the intended domain of science be *everything*? That depends on one's metaphysical presuppositions. I have suggested that theists have reasons for thinking that God cannot be represented with mathematical models in the same way as can the natural world. But in this case, restricting science to the natural world wouldn't be an impediment; in fact, this restriction would *help* science stay focused on a domain that can be mathematically modeled.

Let's turn now to Fales. According to Fales, science aims to "discover the causes of things." But there are many reasons to doubt this claim, casting suspicion on the first premise of his anti-methodological naturalism argument. First of all, as Bertrand Russell pointed out long ago (Russell 1912), the notion of "cause" has disappeared from fundamental physics. And Russell's claim still holds true: if you open any textbook of quantum field theory, you will find no use of the word "cause." But if fundamental physics is not seeking the causes of things, then science doesn't generally seek to discover the causes of things. What's more, causation is still one of the most problematic notions in metaphysics. Should we really define the aims of science in terms of the opaque notion of a cause? (For a recent argument against the world's having a fundamental causal structure, see Norton 2003.)

But setting aside the fact that causation isn't a central notion in fundamental science, there are many causes in which science displays no interest. For example, my desire to complete this caused me to set my alarm for 6:00 a.m. But science isn't aiming to discover this cause (nor do I think it would, even if the NSF had unlimited funding). What science might try to discover are the causes that are *typically* or *generally* associated with early-morning rising. In other words, science doesn't seem to be interested in discovering all causes; rather, science seems to be interested in discovering reliable causal links.

But now, if science aims to discover reliable causal links, then methodological naturalism wouldn't impede science. In particular, a theist might well think that there are no reliable causal links between God and the natural world – in which case, restricting science to the natural world would only help science achieve its aims by keeping it focused on a tractable domain.

Thus, both Monton and Fales make an unwarranted jump from the fact that science is trying to discover truths to the claim that science is trying to discover *all sorts* of truths. But some people, in particular theists, might think that some truths aren't amenable to scientific investigation.

I will illustrate this point by returning to an earlier analogy. Imagine yourself a prehistoric human, whose primary goal is to acquire food. Now suppose that you propose to adopt the agricultural method – to till the soil, plant crops, and tend them. And suppose that I reply:

> This agricultural method would impede me from getting food. While you are out tilling the soil and planting crops, I could be foraging for whatever kind of food I want. While you'll be limited to eating things like corn and wheat, I'll be able to eat anything that I can find. The agricultural method would place arbitrary restrictions on my diet.

It's completely obvious what's wrong with my way of thinking in this scenario. Tilling the soil and planting crops might entail the loss of some opportunities to collect food, but the upside is that it is systematic and reliable. The "shackling science" argument against methodological naturalism makes the same sort of mistake. The scientific method stands to acquiring truth as the agricultural method stands to acquiring food. Just as there are unsystematic ways of acquiring food, so there are unsystematic ways of acquiring truth. And just as the agricultural method systematically generates food, so the scientific method – when it's working well – systematically generates truths about the natural world.

Nonetheless, the agricultural method has limitations: it can't produce every kind of food. Similarly, the scientific method has limitations: it might not be well-tuned for the discovery of every kind of truth. And of all people, a theist is most likely to think that some truths aren't of the right sort to be fit into a scientific account of the world; some truths simply don't fall under general laws, nor can they be accurately represented by means of mathematical models. That's why a theist shouldn't expect to find God in science – because science works by restricting itself to a more manageable kind of fact.

Conclusion

Discussions of naturalism often generate a lot of heat, but rarely much light. One of the main problems with these discussions is the fact that the word "naturalism" has no precise definition, and yet it carries so much emotive content: some people self-identify as naturalists, and others reject naturalism, even though these two groups of people might agree on almost everything else, including how to do science. In this essay, I have discussed two types of naturalism: on the one hand, naturalism can be a synonym for atheism; on the other, naturalism can be a strategy for scientific investigation. I have argued that the latter form of naturalism has little to do with the former. Indeed, methodological naturalism – a strategic narrowing of investigative focus – finds a highly plausible motivation in supernaturalist theism.

Acknowledgments

For helpful feedback, I thank James Anderson, Robert Bishop, Kelly Clark, Andy Crouch, Robbie Hirsch, Jeff Koperski, Alan Love, Rebecca McLaughlin, and especially Del Ratzsch.

References

Bishop, R. (2013). God and Methodological Naturalism in the Scientific Revolution and Beyond. *Perspectives on Science and Christian Faith* 65: 10–23.

Boudry, M., Blancke, S., and Braeckman, J. (2010). How Not to Attack Intelligent Design Creationism: Philosophical Misconceptions about Methodological Naturalism. *Foundations of Science* 15(3): 227–244.

Clark, K.J. (2014). *Religion and the Sciences of Origins: Historical and Contemporary Discussions*. London: Palgrave Macmillan.

Dembski, W.A. (2004). *The Design Revolution: Answering the Toughest Questions about Intelligent Design*. Downers Grove: InterVarsity Press.

de Vries, P. (1986). Naturalism in the Natural Sciences: A Christian Perspective. *Christian Scholar's Review* 15: 388–396.

Fales, E. (2009). Animadversions on *Kitzmiller v. Dover*: Correct Ruling, Flawed Reasoning. Available from: http://infidels.org/library/modern/evan_fales/intelligent-design.html (last accessed July 15, 2015).

Fishman, Y.I. and Boudry, M. (2013). Does Science Presuppose Naturalism (or Anything at All)? *Science & Education* 22(5): 921–949.

Foster, M.B. (1934). The Christian Doctrine of Creation and the Rise of Modern Natural Science. *Mind* 43: 446–468.

Foster, M.B. (1935). Christian Theology and Modern Science of Nature I. *Mind* 44: 439–466.

Foster, M.B. (1936). Christian Theology and Modern Science of Nature II. *Mind* 45: 1–27.

Haught, J.F. (2004). Darwin, Design, and Divine Providence. In *Debating Design: From Darwin to DNA*, edited by M. Ruse and W. Dembski, pp. 229–245. Cambridge: Cambridge University Press.

Hempel, C.G. and Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science* 15: 135–175.

Hooykaas, R. (1972). *Religion and the Rise of Modern Science*. Edinburgh: Scottish Academic Press.

Hooykaas, R. (1987). The Rise of Modern Science: When and Why? *British Journal for the History of Science* 20: 453–473.

Hoyningen-Huene, P. (2013). *Systematicity: The Nature of Science*. Oxford: Oxford University Press.

Klaaren, E.M. (1977). *Religious Origins of Modern Science: Belief in Creation in Seventeenth-Century Thought*. Grand Rapids: W.B. Eerdmans.

Koperski, J. (2008). Two Bad Ways to Attack Intelligent Design and Two Good Ones. *Zygon* 43: 433–449.

Miller, K.B. (2009). The Misguided Attack on Methodological Naturalism. In *For the Rock Record: Geologists on Intelligent Design*, edited by J.S. Schneiderman and W.D. Allmon, pp. 117–140. Oakland: University of California Press.

Monton, B. (2009). *Seeking God in Science: An Atheist Defends Intelligent Design*. Calgary: Broadview Press.

Norton, J.D. (2003). Causation as Folk Science. *Philosophers' Imprint* 3: 1–22. Available from: http://hdl.handle.net/2027/spo.3521354.0003.004 (last accessed July 15, 2015).

Pennock, R.T. (1999). *Tower of Babel: The Evidence against the New Creationism*. Cambridge, MA: MIT Press.

Pennock, R.T. (2011). Can't Philosophers Tell the Difference between Science and Religion?: Demarcation Revisited. *Synthese* 178: 177–206.

Plantinga, A. (1997). Methodological Naturalism. *Perspectives on Science and Christian Faith* 49(3): 143–154.

Plantinga, A. (2009). The Evolutionary Argument against Naturalism: An Initial Statement of the Argument. In *Philosophy after Darwin: Classic and Contemporary Readings*, edited by Michael Ruse, pp. 301–309. Princeton: Princeton University Press.

Ratzsch, D. (2004). Natural Theology, Methodological Naturalism, and "Turtles All the Way Down." *Faith and Philosophy* 21: 436–455.

Rea, M.C. (2002). *World Without Design: The Ontological Consequences of Naturalism*. Oxford: Oxford University Press.

Ruse, M. (2005). Methodological Naturalism Under Attack. *South African Journal of Philosophy* 24: 44–60.

Russell, B. (1912). On the Notion of Cause. *Proceedings of the Aristotelian Society*: 1–26.

Scott, E. (1993). Darwin Prosecuted. *Creation/Evolution* 13 (2).

Scott, E. (2004). *Evolution vs. Creationism: An Introduction*. Oakland: University of California Press.

Sober, E. (2010). Darwin and Naturalism. In *Did Darwin Write the* Origin *Backwards? Philosophical Essays on Darwin's Theory*, edited by Elliot Sober. Amherst: Prometheus Books.

Sober, E. (2011). Why Methodological Naturalism? In *Biological Evolution – Facts and Theories, A Critical Appraisal 150 Years After "The Origin of Species,"* edited by M. Leclerc, G. Aulette, and R. Martinez, pp. 359–378. Rome: Gregorian Biblical Press.

van Fraassen, B.C. (1991). *Quantum Mechanics: An Empiricist View*. Oxford: Oxford University Press.

van Fraassen, B.C. (2002). *The Empirical Stance*. New Haven: Yale University Press.

Watkins, E. (2012). *Kant: Natural Science*. Cambridge: Cambridge University Press.

11

Naturalism and the Question of Realism

DREW KHLENTZOS

Introduction: What Is Naturalism?

This chapter explores the relation between naturalism and realism. Accepting both of these philosophical viewpoints has consequences for how we understand the nature and scope of philosophy. Where realism, the view that the world exists independently of the human mind (that the world is the way it is independent of our conceiving of or thinking about it), commands a broad consensus in philosophy, naturalism sharply divides philosophical opinion. Some philosophers are disinclined to accept it because they worry about the meaning of "naturalism": Is it the idea that we should trust only in the methods of science? A laudable thought, no doubt, but not sufficiently precise to command their assent:

What, after all, are the "methods of science"?
What is to count as science?
Must naturalists foreswear the right to subject the methods of science to philosophical scrutiny?

Such philosophers think these questions have to be answered before we can sensibly be said to believe or disbelieve in naturalism.

Yet if this is a good reason to withhold assent from naturalism, it is equally a reason to withhold assent from science: something most philosophers are loath to do. For the very same questions recur for anyone who accepts science and rejects certain theories as unscientific:

Should we trust in the methods of science?
What are these "methods"?
What, after all, is science?
Must anyone who believes in science and rejects pseudoscience foreswear the right to subject the methods of science to philosophical scrutiny?

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

It is not that these questions are unimportant or that there is nothing to be gained by trying to answer them. On the contrary, philosophy proves its worth in attempting to do so. The point is that either there are answers to these prior questions about science, in which case the proposal that we should trust only in science and scientific methods is a perfectly meaningful one, or there are no such answers to be had, so that neither "naturalism" nor "science" has a determinate sense. If the latter is the case, it needs to be shown how the lack of an adequate definition for "science" can leave science's credibility untouched while a similar lack of a definition for "naturalism" damages naturalism's.

Then what of philosophy? Whatever else it is, philosophy is no science, according to many who have pondered this question. But naturalists beg to differ. They believe philosophy is part of science, if not a science in (or unto) itself. Does philosophy use scientific methods? Most certainly, unless we are to discount its use of logical, probabilistic, statistical, optimizing, and satisficing methods of reasoning. Does it have a domain or methods it can call its own? No, though some philosophers may disagree. Such philosophers identify conceptual analysis either as the domain or the distinctive methodology of philosophy. But neither suggestion seems right. To the extent that their practitioners reflect on the concepts they use, which they often do, all the sciences, not just philosophy, have a place for "conceptual analysis." Moreover, there is every reason for philosophers *qua* philosophers to take an interest in empirical questions: How does human memory work? How does the brain process sound and touch? Are there linguistic universals? How do children acquire a language? These are but a small sample of questions that are of great interest to philosophers: questions that are straightforwardly empirical.

Thus, all naturalists need say about the "problem" of characterizing naturalism is that it is contingent on the prior problem of characterizing science. Perhaps there is more to naturalism than the simple thesis that we should trust only in science and the methods of science, but naturalism is at least that thesis. At any rate, this is all the term "naturalism" need, and for our purposes shall, mean.

The demand for a precise definition of science and its methods should be rejected as a precondition for accepting naturalism. It should be rejected not only because of the selective way it is imposed on naturalists but because such demands are spurious in themselves – instances of the "Socratic fallacy" Peter Geach long ago named and exposed:

> if you know you are correctly predicating a given term "T" you must know what it is to be "T" in the sense of being able to give a general criterion for a thing being T. (1966, 371)

Realism is likewise afflicted with this excessive zeal for definition and spurious or overly exacting demand on how it is to be understood. But even deeper, there is genuine uncertainty as to what "realism" is.

Our interest here lies in the relation between naturalism and realism. Many have thought – I confess, I am one – that naturalistic realism is the most plausible form of realism. But some allege the two are not even compatible. The plan for this chapter is to first, as best we can, characterize both realism and naturalistic realism. We shall then review some arguments, due to Hilary Putnam, designed to show that realism is incoherent, our verdict being that these arguments are not compelling. Then, having argued that naturalistic realism is at least a credible view, we will draw upon the resources of naturalism and realism to cast doubt on a thesis at odds with both: the thesis that philosophical intuitions have a privileged epistemic status, such that unlike the ordinary intuitions of common folk or

those of scientists, philosophers' rational intuitions provide *prima facie* evidence that what is intuited is true.

What Is Realism?

By "realism," I mean metaphysical realism. Hilary Putnam invented this term, investing it with semantic and epistemic content in addition to its more familiar metaphysical and ontological content. We shall look at Putnam's characterization of metaphysical realism, along with some of his well-known arguments against it.

Perhaps the main source of controversy for the characterization of realism is whether realism entails any substantive semantic or epistemic theses. Realists minimally hold that the world and its constituents exist independently of the mind. To be sure, some philosophers reject "mind-independence" as obscure – Rudolf Carnap goes so far as to declare it meaningless:

> The statement asserting the reality of the external world (realism) as well as its negation in various forms, e.g. solipsism and several forms of idealism, in the traditional controversy are pseudo-statements, i.e. devoid of cognitive content. (1963, 868)

Others complain that the mind-independence thesis is coherent but erroneous, since minds themselves are not mind-independent. The relevant sense of "independent" for realism, however, is conceptual or epistemic: minds are "mind-independent" in the sense that the existence of minds does not depend upon any human ability to detect them or to ratify their existence.

Traditionally, realism was opposed to idealism: the view that the world is constituted or constructed by minds or Mind (that reality is ideas or spirit, not matter). That was how things stood until Michael Dummett (1978; 1991; 1993) proposed replacing the mind-independence thesis with (what he took to be) a semantic (near) equivalent: according to Dummett, the realist's assertion of the mind-independence of things in the world is just a metaphorical way of alluding to the semantic fact that the law of bivalence holds for every meaningful statement (apart from those containing vague terms). This law states that every statement is either true or false.

Dummett's "antirealist" contests bivalence. He or she maintains that we are not entitled to assert bivalence for statements for which we have no evidence. An example would be:

(1) Themistocles slept restlessly the night before the Battle of Salamis.

Whereas realists regard (1) as determinately either true or false, depending on whether Themistocles did sleep restlessly the night before the Battle of Salamis, Dummett's antirealist demurs, claiming that the content of an assertion is determined by the evidence that would render it true, not by undetectable states of affairs (such as Themistocles' putative disturbed sleep that fateful night in 480 BCE). As a result, the antirealist refuses to assert bivalence for (1).

Antirealism, thus construed, is not idealism. Though taking their lead from intuitionists, who treat mathematical objects as mental constructs, antirealists need not regard objects in the world as mental constructs. Even so, in Dummett's hands, realism becomes a thesis

about truth and meaning. Many worry that this semantic reformulation misrepresented realism's core metaphysical character: surely, realism is a thesis about what the world is like, not how meaning is determined.

The mind-independence characterization of realism raises a host of semantic and epistemic questions. How could we have reliable knowledge of a mind-independent world? (Kant's question.) If realism is true, how is radical skepticism to be avoided? (Descartes' question.) How is it possible to so much as represent, let alone reliably represent, a mind-independent world in thought or language? (The representation problem.[1]) The debate about whether realism rests upon semantic foundations and the relation between realism and skepticism assumed prominence through the work of Hilary Putnam.

Hilary Putnam on Metaphysical Realism

At one time realism's most eloquent spokesman, Hilary Putnam went on to become its most forceful critic, eventually declaring metaphysical realism to be "incoherent," by which he means not that it is logically inconsistent, but that "the attempts at clear formulation never succeed in capturing the content of 'metaphysical realism' because there is no real content there to be captured" (Putnam 1992b, 353).[2]

The type of metaphysical realism Putnam takes issue with is the naturalistic one he had previously defended; thus, his question is whether metaphysical realism can be given a suitable naturalistic justification. While many, including Putnam, have thought that science attests to the existence of a mind-independent world, Putnam would come to argue that it does not.[3] We shall now review Putnam's two most important (and controversial) arguments designed to show the incoherence of realism: the model-theoretic argument (MTA) and the

1 For discussion of the representation problem for realism, see Khlentzos (2004; 2011).

2 Putnam does not mean to signal thereby that he has embraced either idealism or Dummett-styled antirealism. Indeed, though he had hoped to define the notion of truth in terms of idealized rational acceptability, he has always rejected antirealism about the past: "We certainly regard the statement that Lizzie Borden murdered her parents with an axe as one that has a determinate truth-value (either she did or she didn't administer the famous 'forty whacks'…any theory of truth that makes the truth or falsity of a historical claim depend on whether that claim can be decided in the future is radically mistaken" (1992b, 357).

3 Putnam claims that every attempt to distinguish the substantive claims of metaphysical realism from a kind of scientific realism he and Richard Boyd had formulated and defended fail. On the Putnam–Boyd version of scientific realism, "the success of the mature physical sciences is explained by the fact that the terms used in those sciences typically refer, and that the statements that constitute the basic assumptions of those theories are typically approximately true" (1992b, 354). Putnam remarks, "I saw this as science's explanation of the success of science, rather than as a metaphysical explanation of the success of science" (1992b, 353).

The problem with this type of scientific realism, he argues, is that the explanation it provides of the success of science is perfectly compatible with metaphysical antirealism. That seems quite correct. There is nothing in the Putnam–Boyd argument for scientific realism that antirealists cannot accept. The argument says nothing about the metaphysical nature of the entities science studies and posits, only that scientific terms typically refer and scientific theses are typically approximately true. Provided the notion of truth is a humanly recognizable one, antirealists can happily concur. If the only way in which metaphysical realism can be justified naturalistically is by some such semantic argument based on the success of science, then Putnam's claim that realism is incoherent is no longer so puzzling: the idea is, everything the realist wishes to say, antirealists can endorse, so that realism becomes ineffable. However, some naturalistic-minded philosophers see this as an intrinsic shortcoming of semantic defenses of realism, rather than a demonstration of the ineffability of realism.

brains-in-a-vat argument (BIVA). Since the latter argument is easier to comprehend than the former, we shall begin with it.

The Brains-in-a-Vat Argument

Putnam believes metaphysical realism engenders radical skepticism. He is not alone in thinking so. The skeptic first appears in ancient philosophy, and, ever since Descartes conceived of his Evil Demon, realists have struggled to answer the skeptic's challenge: If the world is independent of the mind, as realists assert, then how do we know that all our beliefs about that world are not mistaken?

The BIVA scenario is simply a modern reworking of Descartes' Evil Demon. Namely, for all we could ever tell, we could be brains in a vat: brains kept alive in a bath of nutrients by alien scientists. All our thoughts, all our experience, all our science would be systematically mistaken if we were. We'd have no bodies, though we'd think we did; our vat world would contain no physical objects, yet it would seem to us that it did; there'd be no earth, no sun, no vast universe in our vat world, only the brain's deluded representations of such. At least, this could be the case if our representations derive their content from links with mind-independent objects and states of affairs. Since realism implies that this absurd possibility could hold without our being able to detect it, it has to be rejected.

Putnam's BIVA introduces something new to the skeptic's challenge. BIVA purports to show realism is incoherent, rather than unwarranted or false. The reason, according to Putnam, is that realism must concede the possibility of our being brains in a vat, yet it entails something inconsistent with that possibility: namely, that if we were to be so envatted, we could not have the thought that we were. A crucial assumption of the argument is semantic externalism: the thesis that the reference of our words and mental symbols is partially determined by contingent relations between thinkers and the world. This is a semantic assumption that many realists independently endorse. Thus, semantic externalists hold, for example, that in a world otherwise indistinguishable from ours that lacked any water but instead had a substance that filled that world's oceans, lakes, and rivers and fell as rain, a substance indistinguishable from real water in appearance, taste, and the functions people in that world used it for, that substance would not in reality be water even if it were to go by the name "water." "Water" refers to the substance whose chemical composition is H_2O. On this "twin" earth, twin water (twater) is not composed of H_2O but of some other stuff: XYZ, say. Hence, twater is not water, according to semantic externalists. So our term "water" refers to water partly because of a causal link between it and the substance H_2O. Twin-earthers' homophonic term "water," on the other hand, refers to twater rather than water (i.e., the substance composed of XYZ rather than H_2O) partly because of the causal link between their use of "water" and the substance XYZ.

Given semantic externalism, Putnam argues that we can know *a priori* that we are not brains in a vat, a possibility we could not represent, let alone know about, if realism were true. In essence, his argument is this:[4]

(1) A brain-in-a-vat's word "brain" does not refer to brains.
(2) My word "brain" refers to brains.
(3) ∴ I am not a brain in a vat.

4 This formulation is Tim Button's (2013, 118).

Semantic externalism justifies (1), since a brain-in-a-vat's thoughts are merely electrical signals emanating from its brain. The principle of disquotation justifies (2).[5] But can Descartes' skeptic really be silenced in such a summary fashion? There is reason to doubt whether or not Putnam has proven we can know *a priori* that we are not brains in vats rather than simply implicitly assumed that we are not. For, suppose I am wondering whether I am a brain in a vat and assume for the sake of argument that I am. I read over Putnam's proof and think, "But if I *am* a brain in a vat, (1) and (2) seem to pull in opposite directions: either 'brains' are just the vat's electrical signals that I mistake for genuine human brains or they are human brains. If 'brains' are human brains then, while (1) is true, (2) is false. If 'brains' are just the vat's electrical signals then, while (2) is true, (1) is false." I am led to conclude that the BIVA proof that we can know *a priori* that we are not brains in a vat appears unsound.

Perhaps Putnam's proof that we can know that we are not brains in a vat is unsound. Nonetheless, the challenge for the realist remains if Putnam's analysis is correct: realists are committed to the possibility of our being brains in a vat, yet, given semantic externalism, such a possibility is unthinkable.

Realists have a principled reason to question Putnam's analysis. For realists, possible states of affairs are no more determined by our intuitions than are actual states of affairs. To say that it is possible that we are brains in a vat is to say that there is a possible world in which we are. Is there such a possible world? Could our world be a brains-in-a-vat world? We don't know. Just because we can't see a contradiction in supposing it could is no guarantee that it could. Even if it seems to us that the brains-in-a-vat scenario is possible, it may not be (it seems possible that every even number is the sum of two prime numbers, but it might, in fact, be impossible). Hence, realists should say of the brains-in-a-vat hypothesis that it is *at best* **conceivable** that we are brains in a vat. It is the skeptic's task to go on to prove that the brains-in-a-vat scenario is genuinely (i.e., metaphysically) possible.

The Model-Theoretic Argument

Putnam's MTA uses technical results in mathematical logic (we shall not rehearse these technicalities) to argue that it is incoherent to suppose that an ideal theory of the world could fail to be true. Since realism is committed to the thesis that an ideal theory of the world could be false, realism is refuted if the MTA is sound.

It might help to first present some of the key technical ideas informally.

Consider three runners, Alice, Billie, and Claire, vying for supremacy over 100 m. Suppose Alice is faster than either Billie or Claire and that Billie is faster than Claire. Suppose also that Claire is taller than either Alice or Billie and Alice is taller than Billie. In logic, a theory is just a collection of sentences. So consider the theory consisting of our six sentences:

(1) Alice is faster than Billie.
(2) Alice is faster than Claire.
(3) Billie is faster than Claire.

5 The principle of disquotation for reference allows one to infer what a term in quotation marks refers to by removing the quotation marks, as in the base clauses of a recursive definition of truth such as those Alfred Tarski used to define truth: (1) "snow" refers to snow; (2) x satisfies "is white" if and only if x is white. The name "disquotation" seems to be due to Quine.

(4) Claire is taller than Alice.
(5) Claire is taller than Billie.
(6) Alice is taller than Billie.

Suppose we let the letter a stand for Alice, b stand for Billie, and c stand for Claire. These letters are called individual constants. Let us use the notation "Fxy" to represent "x is faster than y" and "Txy" to represent "x is taller than y". The letters F and T are called predicate letters. We can represent the theory in (1)–(6) using this notation as follows:

(i) Fab; (ii) Fac; (iii) Fbc; (iv) Tca; (v) Tcb; (vi) Tab.

Now the theory talks about three runners, Alice, Billie, and Claire. When we represent the theory formally by sentences (i)–(vi), we intend it to describe those three runners and we want what it says about them to be true if it is true. In logic, we distinguish theories (which are sentences) from the things they talk about (which are generally not sentences). The things the theory talks about feature in abstract entities called **structures** as the **domains** of the structures. What the theory says about these things receives an **interpretation** in the structures.

An **interpretation function** assigns objects to individual constants, pairs of objects to binary relations (e.g., "faster than" and "taller than" in our example). Even when the domain is the one specified in the theory, there are structures that don't render the theory true. Thus, for a structure *S* in which a refers to Claire, b to Billie, and c to Alice, where F is now the relation of being taller than and T is the relation of being faster than, as these relations were specified in (1)–(6), sentences (i)–(vi) will not all come out true:

(i) Fab will come out true in *S*, since Claire is taller than Billie.
(ii) Fac will come out true in *S*, since Claire is taller than Alice.
(iii) Fbc will come out *false* in *S*, since Billie is not taller than Alice.
(iv) Tca will come out true in *S*, since Alice is faster than Claire.
(v) Tcb will come out true in *S*, since Alice is faster than Billie.
(vi) Tab will come out *false* in *S*, since Claire is not faster than Billie.

A structure S^* with domain {Alice, Billie, Claire} that assigns Alice to a, Billie to b, and Claire to c and interprets the relations *faster than* and *taller than* in the way the theory specifies will render all of (i)–(vi) true; that is, make the theory come out true. Such a structure that makes all of a theory's sentences come out true is called a **model** of the theory.

A model of a theory renders a theory true by making all of its sentences true. But that does not mean that the model must interpret the theory correctly. We've seen a structure, namely *S*, that does not interpret the theory in the right way. Not all of the theory's sentences come out true in *S*, so *S* is not a model of our theory. But there are structures that interpret the theory in a deviant way which nonetheless make all of the theory's claims come out true. Such models differ from S^* in their interpretation of the theory. Other models may differ from S^* in their domains, as we shall see further on.

Consider a structure $S^\#$ with domain consisting of the numbers 1, 2, and 3, in which the individual constant a refers to 1, b to 2, and c to 3, wherein predicate F is interpreted as the relation of *being less than* (<), and predicate T as the relation of *being not equal to* (≠). $S^\#$ is a model of our theory, because it renders (i)–(vi) all true: Fab asserts in $S^\#$ that $1 < 2$,

Fac asserts that 1 < 3, Fbc asserts that 2 < 3, Tca asserts that 3 ≠ 1, Tcb asserts that 3 ≠ 2, and Tab asserts that 1 ≠ 2. All these statements are true statements about the numbers 1, 2, and 3. But clearly $S^{\#}$ is not the model the theory is describing. Our theory makes claims about female athletes, not numbers. Since $S^{\star}$ interprets the theory correctly, $S^{\star}$ is the **intended model** of the theory. In contrast, $S^{\#}$ is clearly *not* the intended model. The notion of an intended model can be used to connect truth for a formalized theory with truth for the theory it represents: a formalized theory is deemed **true** if it is **true in its intended model**.

We represented our theory about Alice, Billie, and Claire in a very simple formal language that used individual constants a, b, c and predicate letters F, T. A richer version of this logic, known as first-order logic, contains in its language not only these symbols but also symbols for logical operators, such as negation (~), conjunction (&), disjunction (v), the conditional (→), and equivalence (↔), as well as for expressions of generality (known as quantifiers), such as "all" (∀) and "some" (∃), together with individual variables x, y, z,…

In 1930, Kurt Gödel proved that every consistent set of first-order-logic sentences has a model. Gödel's completeness theorem, as it is known, is crucial to the MTA. Putnam used the completeness theorem (and other logical theorems) to argue that an ideal theory must be true. His argument can be represented thus:

(1) Suppose we had an ideal theory **T** that passed every observational and theoretical test we could conceive of. Assume **T** can be formalized in first-order logic. Assume also that the world is infinite in size and that our ideal theory **T** says it is. Assume, finally, **T** is consistent.
(2) Then, as **T** is a consistent first-order-logic theory, by the completeness theorem for first-order logic, **T** will have a model.
(3) Then, by another logical theorem, there exists a model of **T** that is the same size as the world. Call this model *M*.
(4) Nothing in the construction of *M* guarantees that the objects in its domain are objects in the real world. The domain of *M* may consist wholly of numbers, for example, just as the domain of $S^{\#}$ did in our previous example.
(5) But a logical theorem can be used to generate from *M* a new model *W* whose domain does consist of the objects in the world and which assigns to all the predicates of **T** subclasses of its domain and relations defined on that domain.
(6) We now have a correspondence between the expressions of the language *L* in which *T* is expressed and (sets of) objects in the world, just as the realist requires.
(7) **T** will then be true if "true" amounts to "true-in-*W*."
(8) If **T** is not guaranteed true by this procedure, it can only be because *W* is not the *intended* model.
(9) Yet all our observation sentences come out true according to *W*, and the theoretical constraints are satisfied, because **T**'s theses all come out true in *W* also. So *W* surely is an intended model.
(10) Hence, we can conclude **T** is indeed **true**, since a theory is **true** if it is true in its intended model, and *W* is an intended model.

At points (3) and (5) of this argument, reference was made to a logical theorem that allowed one first to ensure that the model of **T** would be of the same size as the world and then that this model could be projected on to a model that contained elements from the world. The

logical theorem concerned is a generalization of the Löwenheim–Skolem theorem (LST), which states that any consistent, countable set of first-order formulae has a countable model: in fact, a model in the set of integers I; that is, the set {0, 1, 2,… }.

A "countable" model, as the name suggests, is a model whose domain consists of elements that can be counted. The distinction between things that are countable and things that are not is reflected both in ordinary perceptual experience and in language by the use of mass terms as opposed to count nouns: we can count the ducks on the pond but we cannot count the water in the pond: "water" is a mass term but "duck" is a count noun. "How many ducks are there on the pond?" is a meaningful question; "How many *waters are there in the pond?" is not. In 1874, the German mathematician Georg Cantor proved that the number of points on a continuous line is uncountable. The points on a continuous line correspond to what are known as real numbers. Uncountable sets such as the set of real numbers, $\mathfrak{R}$, are also called nondenumerable sets. Cantor's work in set theory provided the foundation for all subsequent work. Set theorists such as Ernst Zermelo, Abraham Fraenkel, Thoraf Skolem, and Hermann Weyl built on Cantor's work in an effort to find axioms for set theory. The eventual result was known as Zermelo–Fraenkel (ZF) set theory.

The LST has a surprising consequence, called Skolem's paradox: Since the axioms of ZF set theory comprise a consistent, countable set of first-order formulae, by the LST, the ZF axioms have a model in the set of integers. Yet one can prove in ZF that nondenumerable sets exist, such as set $\mathfrak{R}$.

Recall from our previous discussion that a model of a theory is a structure that makes all of the theory's sentences come out true. Thus, even though ZF has a theorem that states "Uncountable sets such as the set $\mathfrak{R}$ of real numbers exist," there are models of ZF whose domains consist solely of the integers {0, 1, 2,… }; that is, models that only have a countable number of elements in their domains. How, then, can ZF's theorem about the existence of uncountable sets such as $\mathfrak{R}$ possibly come out true in a structure such as the set of integers {0, 1, 2,… } whose domain contains *no* uncountable elements whatsoever? This is Skolem's paradox.

In his MTA, Putnam argues that, since *W* passes every conceivable operational and theoretical constraint, *W* is the intended model of **T**. **T** must then be true, Putnam claims, since truth is truth in the intended model.[6]

However, realists such as Lewis (1983; 1984) and Devitt (1983) object that nothing in *W*'s construction guarantees that **T**'s terms stand in the right reference relation to their

6 Putnam has a fallback argument if the argument based on the LST fails. This is his permutation argument: Suppose the realist is able to somehow specify the intended model. Call this model W^*. Then nothing the realist can do can possibly distinguish W^* from a permuted variant *P*, which can be specified following Putnam (1994b, 356–357). We define properties of being a cat* and being a mat* such that:

I. In the actual world, cherries are cats* and trees are mats*.
II. In every possible world, the two sentences "A cat is on a mat" and "A cat* is on a mat*" have precisely the same truth value.

Instead of considering the two sentences "A cat is on a mat" and "A cat* is on a mat*," now consider only the one: "A cat is on a mat." Allow its interpretation to change by first adopting its standard interpretation and then adopting the nonstandard interpretation in which the set of cats* is assigned to "cat" in every possible world and the set of mats* is assigned to "mat" in every possible world. The result will be that the truth value of "A cat is on a mat" does not change and is exactly the same as before in every possible world. Similar nonstandard reference assignments can be constructed for all the predicates of a language (see Putnam 1985; 1994b).

referents in *W* – and justly so, if the "guarantees" in the form of theoretical constraints on reference "themselves are revised in the reinterpretation," as Quine (1982) puts it in the context of discussing Skolem's paradox.[7] To this, Putnam replies that a constraint on reference is "just more theory": if it is a legitimate constraint on **T** at all, *W* will already satisfy it, along with all the other theoretical constraints of **T**, since *W* makes all of **T**'s theses true.

Putnam's "just more theory" gambit shows that any constraint can be so interpreted as to come out true in *W*. But that is precisely the difficulty Lewis (1983; 1984) and Devitt (1983) identified in the first place: *every* model of **T**, even those that do *not* satisfy the right reference constraint, will make all of **T**'s theses true, including **T**'s assertion that the right reference constraint is satisfied *even when it is not*. Thus, one can no more infer that *W* satisfies the right reference constraint because it makes true a sentence expressing the right reference constraint than one can infer that a countable model of ZF contains set $\mathfrak{R}$ because it makes true a ZF sentence stating that $\mathfrak{R}$ exists. As a result, the MTA fails, being based on a mistaken inference.[8]

The provisional conclusion to be drawn from our review of Putnam's BIVA and MTA against realism is that naturalism and realism seem compatible, and that *pace* Putnam, metaphysical realism is not incoherent. To be sure, realism faces difficult challenges: to explain how mental representation of a mind-independent world is possible and to explain how knowledge of such a world is to be had; challenges that are still unmet. However, naturalistic realism offers us a compelling picture of a world not of our making that is often at odds with our thinking, a world revealed by hard-won methods of discovery. Nonetheless, if naturalistic realism is true, then the nature and function of philosophy could not be as many philosophers see them to be. For many take philosophy to be an area of inquiry whose questions are to be resolved purely by reflecting on our concepts or by other forms of *a priori* reasoning. Such theorists believe philosophy is a domain in which *a priori* intuitions not only provide the means for discovering substantive metaphysical, epistemological, or moral truths, but also afford the primary source of evidence that what is intuited is true. However, naturalistic realists cannot accept this view of philosophy or philosophical method. For without a well-confirmed scientific theory ratifying the reliability of *a priori* intuitions, the presumption that they are reliable is baseless according to naturalism, and the assumption that *any* proposition is likely to be true merely because it is intuitively plausible is flatly inconsistent with realism. So, if naturalistic realism is true, a widely accepted view of philosophy and its methods will have to be abandoned. It is time to look more closely at that view.

7 Quine is discussing the discrepancy Skolem's paradox brings to light between what is provable in a theory and what can be verified by a theory's models. The full quotation from which this line is drawn is this: "This consequence has been viewed as paradoxical, in the light of Cantor's proof that the real numbers cannot be exhaustively correlated with integers. But the air of paradox may be dispelled by this reflection: whatever disparities between the real numbers and integers may be guaranteed in those original truths about real numbers, the guarantees themselves are *revised* in the *reinterpretation*" (Quine 1982, 210).

8 Having insisted on the distinction between a model's genuinely satisfying a constraint and the model merely making true a sentence stating the constraint, Lewis (1984) puzzlingly goes on to argue that more is needed to answer Putnam's MTA. The impression persists to this day that the MTA somehow forces the realist to come up with a viable account of reference determination on pain of refutation. But if the MTA is invalid, as I have argued, then there is no reason for the realist to even try to articulate those constraints that tie down reference. Indeed, for all the realist *qua* "realist" knows or cares, the actual determinants of reference might forever elude our grasp.

Naturalism and the Status of Rational Intuitions

Earlier in this chapter, we characterized philosophy as a science without a distinctive domain or methods. From the naturalistic perspective, philosophers' intuitions enjoy no epistemic privileges in comparison with those of psychologists or linguists or physicists. These disciplines regard a theorist's intuitions as nothing more than hunches about how the facts might go, guesses which need to be empirically tested in the form of hypotheses. For linguists, neuroscientists, biologists, or other scientists, intuitions per se have no scientific value. Indeed, a scientist may have conflicting intuitions about a question or even no intuitions at all about it before setting out to test a hypothesis about the right answer to that question. What matters for science is the hypothesis, not the hunch that spawned it. Moreover, in the case of mathematics – an *a priori* discipline, like philosophy – intuitions count for naught: proof is all that is important.

How Could Philosophical Intuitions Be So Different?

Many philosophers vigorously contest the naturalist's attempt to assimilate philosophical intuitions to the intuitions of scientists. For them, the philosopher's rational intuitions, unlike the scientist's empirical ones, do indeed enjoy a special epistemic status. Anna-Sara Malmgren (2011) gives eloquent voice to this view:

> What demarcates philosophy from other academic disciplines, specifically from (other) sciences? One striking difference is that in philosophy we typically do not subject our hypotheses and theories to empirical testing – somehow it is supposed to be sufficient to test a theory in *thought*…the philosopher sits back and runs a thought experiment. How could that be enough? Indeed, how could an experiment performed in thought tell us *anything* about the nature of knowledge, consciousness, time, moral value, or any of the other things philosophers are interested in?
>
> To get clear on this, we first of all need a model of the relevant test procedure – the thought experiment, as used in philosophy. In brief outline, it has the following structure: the hypothesis or theory that is under evaluation states or entails some modal claim (typically a necessary bi-conditional or one-way implication) and in a thought experiment we check that modal claim against our intuitive verdict on an imaginary problem case. If the claim conflicts with our intuitive verdict, this is treated as strong evidence against the theory – indeed, the theory may be abandoned as a result. We say that we have found *a counterexample* to it. If not, this is treated as at least some evidence in support of the theory. We say that it *accommodates our intuitions* about the case. (2011, 263–264)

Malmgren defends the rationalist model of intuition just sketched against the objections of Timothy Williamson, who argues that intuitive judgments are simply judgments about certain sorts of counterfactual conditionals, and that as a consequence we do not need to appeal to a special faculty of rational intuition to explain how we arrive at them – our general capacity to process subjunctive conditionals suffices to explain this.

The first question naturalists will want to ask concerns the evidence for the existence of a special faculty of rational intuition. The second concerns its reliability. As regards the first question, the rationalist presumably does not think that everyone has the capacity to detect the modal claims entailed by a given philosophical hypothesis. Such a faculty must be

developed and refined through philosophical training, even if the general capacity for rational assessment of possibilities is there in all of us, as it presumably is as a prerequisite for rational choice. Philosophical education supplies the concepts, together with the techniques necessary to train our general proclivity for intuitive judgments about possibilities, so that this faculty can then detect the modal consequences of various philosophical hypotheses.

If this is how we come by our rational intuitions, there is no need to posit any specialized subfaculty to explain how we are able to detect the modal consequences of philosophical hypotheses: just as the expert chess player can see that a certain possible move is inconsistent with White's winning the game, whereas the novice she's instructing cannot, so the philosopher's "expert intuitions" about modal consequences reflect her specialized training as a philosopher. In short, learning the language of philosophy and acquiring the relevant philosophical concepts and techniques is all that creatures already adept at conceiving of possibilities need in order to detect the modal consequences of philosophical hypotheses.

Rationalists will not be impressed by this deflationary account of our modal intuitions. Of course we need training in philosophical reasoning before we are able to detect the relevant possibilities. To that extent, philosophers are no different from scientists who need special training in the use of electron microscopes before they can detect molecules or in the use of radio telescopes before they can detect distant moons. But what needs explaining is how, merely by thinking about the matter, philosophers can arrive at reliable conclusions, not just about how the world actually is – the scientist's concern – but also about how the world *must* (or *could not*) be. Thus, merely by reflecting on the nature of computation, John Searle argued in his Chinese room thought experiment that the mind could not be any sort of computational device, and, likewise, David Chalmers contended that our ability to clearly conceive of physical duplicates of ourselves who lack any consciousness could be used to demonstrate that materialism must be false. Surely these examples and many others like them show that we do indeed have a faculty of rational intuition that allows us to detect metaphysical possibilities?

In fact, neither Searle's nor Chalmers' thought experiments enjoy a wide consensus among philosophers: both the Chinese room and zombie arguments are regarded as contentious, and in both cases it is precisely the rational intuition of a certain metaphysical possibility that is contested. Thus, rational intuitions differ from one philosopher to another, depending on the theory to which each happens to subscribe. Moreover, it is not as though the Chinese room argument or the zombie argument are outliers in this regard: philosophical thought experiments are regularly contested by philosophers who either do not share the rational intuitions they are designed to elicit or else question the validity or relevance of those intuitions, if they do.

If rational intuitions are, in general, theory-dependent, the attempt to assimilate those intuitions to a set of data or to the observations a philosophical theory must account for is thus fraught with difficulty. So, the rationalist must reject the idea that intuitions are theory-dependent. For otherwise, intuitions per se would have no evidential value: whether they were reliable would depend on whether the theory that grounded them was true, and thus there would no longer be any presumption that rational intuitions are sound judgments about abstract metaphysical possibilities, with the result that our access to modal facts would be no more direct than our access to nonmodal facts.

The rationalist could reject the theory-dependence of intuition by arguing that our intuitions derive from our grasp of the relevant concepts. If all that is required to discern a

metaphysical possibility is *a priori* reasoning from our own concepts, then any support this or that theory provides for those intuitions is inessential: all theories do is flesh out the intuitions and place them in a wider explanatory context. But one does not need a theory to intuit that a certain state of affairs is possible.

Are Modal Intuitions Reliable?

For rationalists, it is an article of faith that our modal intuitions are reliable. But are they in fact? Could we just be mistaken in judging something to be possible when it really isn't, or in judging something impossible when in fact it is possible? Could we feel completely confident in our judgment that something is possible/impossible though it is in fact impossible/possible?

The last half-century of research into human reasoning has shown not only that we regularly and systematically err in our assessments of what is possible and probable, but also that these erroneous judgments are often associated with levels of confidence just as high as any associated with sound judgments. Where it has been possible to compare our judgments with the relevant external facts, psychologists have been able to determine how well-calibrated those judgments are. Where this has not been possible, as in our estimates of probability, subjects' judgments have been compared with normatively correct verdicts instead. The surprising finding is that our judgments are often most poorly calibrated with the facts when we are most (or least) confident.[9]

It might be thought that these findings have little relevance to the question of whether philosophers' judgments about metaphysical possibilities are reliable, since in this domain philosophers have an expertise that ordinary folk lack. Naturalists agree that philosophers have such an expertise, but the question is whether this expertise is best conceived as a proficiency in generating accurate judgments about metaphysical possibilities or merely as a capacity to generate such judgments in the first place. Naturalists surmise it is the latter. That does not commit them to regarding the philosopher's modal judgments as being just as likely to be false as true.

The same research tradition that revealed a lack of correlation between the reliability of our judgments and the levels of confidence we invest in them has provided an insight into the nature of intuitive judgments. These, according to the "heuristics and biases" tradition, typically arise in consciousness without any clue as to their origin, being frequently experienced as "gut feelings," which are often associated with high levels of subjective certainty.[10]

9 See, for instance, Oskamp (1965), Fischoff, Slovic, and Lichtenstein (1977), Lichtenstein and Fischoff (1977), Lichtenstein, Fischoff, and Phillips (1982), Dunning et al. (1990), and Vallone et al. (1990).

10 Daniel Kahneman provides a nice illustration of how one such heuristic works. The heuristic detects incongruity by automatically and rapidly checking information against stereotypes without our being aware of it or our having decided to implement it: "We can detect incongruity and abnormality, without thinking, and very, very rapidly. And the subtlety of it is amazing. To give you an example, my favourite: you have an upper-class British voice, a male British voice, saying 'I have tattoos, I have large tattoos all down my back.' In 250 milliseconds the brain has a response: something is wrong. People who speak in upper-class British voices don't have large tattoos down their back…The world knowledge that is brought to bear is enormous – the amount of information that you would have to know and be able to bring to bear instantly in order to know there is something incongruous here – and that happens at a speed that is extraordinary. This is just about as fast as anything in the brain – between 200 and 400 milliseconds there is a full-blown surprise response" (Brockman 2011, 398).

Thus, subjects asked to estimate how likely it is that Fred has a form of cancer that on average strikes 1 in 200, given he has just returned a positive test for it, will confidently answer "Around 90%" when told the test is 98% accurate (the correct answer is around 20%). The tendency to engage in base rate neglect, as Tversky and Kahneman (1982) call it, isn't extinguished by information about the base rate, or even when the base rate is presented in full focus.[11] Furthermore, our probability estimates can sometimes be skewed by deeply entrenched false beliefs of which we have no cognizance.[12]

It is implausible that our estimates of probability should be unreliable but our estimates of possibility reliable. But perhaps the philosopher, through her training, is expert at detecting metaphysical possibilities. Surely, once the philosopher knows the normative theory, she will see through the illusions that trap ordinary folk? Perhaps. But which normative theory of modality tells us whether zombies are metaphysically possible? Moreover, even if expertise is relevant here, along with expertise come various types of illusion of expertise. In his *Thinking Fast and Slow*, Daniel Kahneman comments on the success of investment advisors at predicting the stock market:

> the subjective experience of traders is that they are making sensible educated guesses in a situation of great uncertainty and…the people who pick stocks are exercising high-level skills…that require extensive training. (2012, 215)

Yet in spite of this, "professional investors failed a basic test of skill: persistent achievement" (Kahneman 2012, 214). Success at predicting the movements of shares on the stock market might appear worlds away from detecting whether some state of affairs or situation is or is

11 In their famous 1973 "Tom W." experiment, Tversky and Kahneman divided subjects into three groups assigned the task of estimating the proportion of US graduates in nine different areas of study. Group 1 simply estimated the proportions, but Group 2 and Group 3 were given additional tasks that centered on a fictitious character called Tom W., described as "highly intelligent, neat, orderly, socially inept, self-centered but highly ethical." Group 2 was asked to judge how closely Tom W. resembled a typical graduate student from each of the nine fields. Group 3 were asked to rank the nine fields of study according to the probability of Tom being in them. The results showed Group 1 had estimated the proportions pretty accurately: computer science was identified as the second smallest group, at 7%; social work the second largest, at 17%. As to the similarity question for Group 2, Tom W. was judged most like a computer science graduate. Remarkably, Group 3 members, having judged computer science graduates less numerous than social work graduates, went on to judge it more likely Tom W. was among the computer science graduates than that he was a social work graduate!

12 An oft-replicated experiment in cognitive psychology involves sitting a subject in front of two lights – one red, one blue – and asking them to predict which of the lights will flash on each trial. In fact, as subjects quickly discover for themselves, the lights have been programmed to flash randomly within the constraint that red flashes 70% of the time and blue 30%. What color should a subject predict at each trial? The interesting result is that even though subjects soon discover red flashes 70%, blue 30% of the time, they try to figure out the underlying mechanism determining which light will flash at each trial, switching from red to blue and back again in an effort to "second-guess" the random flashes, with the result that they correctly guess red around 70% of the time and blue around 30% of the time. They thus try to get the prediction for each individual trial correct, resulting in suboptimal accuracy of around 58% overall (in 100 trials, 70 on average will be red and 30 will be blue; given their red accuracy is 70% and their blue is 30%, subjects should correctly guess 49 red, 9 blue). Underlying their error is a misplaced desire for certainty, coupled with the illusion that it is attainable. If they gave up the impossible aim of accurately predicting the outcome of each trial and realized they must accept error in order to reduce it, they would also come to realize that the rational strategy is to predict red at each trial, with a resulting 70% success rate. The mistaken tendency persists even when subjects are offered financial or other incentives for the number of guesses they get correct.

not possible, but the question of success is the same for both cases: What evidence is there to show that the intuitions of philosophers about what is possible reliably track the truth?

In fact, there isn't any empirical evidence to show this, nor could there be. We cannot evaluate the modal facts independently of our beliefs about them, and there is no normative theory to appeal to for independent verdicts on problematic modal cases. In light of this, we should accept intuitive judgments about metaphysical possibilities for what they are: beliefs about the modal facts. *Qua* beliefs, our modal intuitions have no evidential value in themselves. This does not make philosophical thought experiments worthless. To the contrary, such exercises reliably elicit modal judgments, verdicts that may indeed be true. In light of that, thought experiments have real value.

Rationalists will resist this assessment. They will argue that our intuitions about possibility derive from our mastery of the concepts associated with them. Thus, it might be that the zombie intuition results from our grasp of the concept of consciousness, as David Chalmers argues, or that the Chinese room intuition that cognition cannot be computation derives from our grasp of the concept of meaning, as John Searle argues.

Suppose rationalists are right about this and that all our modal intuitions can be traced back to our mastery of concepts. This would not show that our modal intuitions are not beliefs about the modal facts. Yet it might suffice to confer some special evidential status on those intuitive beliefs: for we surely grasp our own concepts, so it is plausible to think we must be reliable about whether those concepts apply or fail to apply across all possible circumstances – isn't that just what it is to have a concept in the first place?

There are two difficulties with the rationalist's proposal. The first is that the type of possibility now being mooted is *conceptual* possibility rather than, as required, *metaphysical* possibility, so we're owed an argument to show one can reliably infer the second from the first. The second problem is that the rationalist's inference from mastering a concept C to being reliable about whether C applies or fails to apply in all possible circumstances is simply fallacious.

The relation between conceptual and metaphysical possibility is highly controversial. Indeed, it is not just naturalistic-minded philosophers who question the move from the one type of possibility to the other. Scott Soames shows that it is implicit in Saul Kripke's (1980) main argument for the existence of necessary *a posteriori* and contingent *a priori* truths that metaphysical possibility cannot be identified with conceptual possibility (Soames 2006). How could evidence that there is no conceptual contradiction in supposing that p holds be converted into evidence that the world really could be in a p-state?

The problem is that where metaphysical possibility concerns those states the world could occupy, conceptual possibility has to do with how we represent those states in thought. Unless we have some guarantee that the states the world can be in cannot outrun those we can conceive it to be in, we can have no assurance that what is *a priori* possible conceptually is possible in reality.

Naturalistic realists aver that there are principled reasons why there can be no such guarantee: where the alethic metaphysical modalities "necessary" and "contingent" partition the class of *all* propositions, the epistemic (conceptual) modalities "*a priori*" and "*a posteriori*" partition the class of *knowable* propositions only. Since only truths are knowable, to infer that a proposition p is metaphysically possible because it is *a priori* (conceptually) possible requires one first to assume that *all truths are knowable*. But this assumption is inconsistent with metaphysical realism, hence inconsistent with naturalistic realism.

But suppose this general distinction between truth and knowability does not hold for the case at hand, suppose that modal facts are somehow special in this regard: modal truths at least are knowable in principle. Suppose further that this is because there is no viable distinction between metaphysical and conceptual modality. Would it follow that our modal intuitions were a reliable guide to what is and is not possible?

Not unless mastery of the concept(s) associated with an *a priori* modal judgment automatically confers upon a thinker the ability to determine whether those concepts apply or fail to apply in any and every conceivable circumstance. Otherwise, we might judge that C applies in some imagined situation S when in fact it does not, or, alternatively, we might judge that C does not apply in S (or fail to judge C applies in S) when in fact it does.

The idea that we have the ability to determine whether a given concept C either applies or fails to apply in any and every situation S we can conceive of is fanciful. To begin with, concepts are not in general transparent (as required for this hypothetical ability to function in the way envisaged). Concepts that might appear transparent may nonetheless be so rich in meaning that codification of their analytic consequences is not even possible. That this is true of arithmetical concepts is the surprising moral of Gödel's incompleteness theorem.

Second, simple concepts can interact in complex and unforeseen ways. Consider logical concepts. Most speakers master the words "not" and "and" (or the concepts of negation and conjunction) in their language and regularly make inferences involving negation and conjunction in thought as well as speech. So consider the following question, to be answered using only your concepts of negation and conjunction:

Is it more likely that (i) you won't rest well tonight OR (ii) you won't rest well tonight without discovering a cure for cancer in the morning?

If you are like most speakers, you'll have immediately judged (i) far more likely than (ii), and you will have judged it thus because you know your chances of discovering a cure for cancer are zero.

In fact, (i) cannot be more likely than (ii) without violating basic logical principles (at least, if (ii) is interpreted using negation and conjunction). For (i) can be written as ~R, with R being "you will rest well tonight" and ~ representing "not," while (ii) can be written as ~(R & ~C), with C being "you will discover a cure for cancer tomorrow morning." By De Morgan's law, ~(R & ~C) is equivalent to a disjunction, ~R v C. The reason that (i) cannot be more likely than (ii) is now evident: (i) is ~R, (ii) is ~R v C, and a disjunction cannot be less likely than one of its disjuncts.

It is not the fact that we mishandle these probabilities that should cause concern for rationalists. We've already seen that our performance with probabilities is poor. The real worry is why (i) should *appear* so obviously more likely than (ii). For our intuitive judgment is based on an entirely spurious ground: that we have no chance of discovering a cure for cancer tomorrow. True, but entirely irrelevant to the question. Why does it seem otherwise?

The answer appears to be that when we consider (ii), "I won't rest well tonight without discovering a cure for cancer tomorrow morning," ~(R & ~C), only one possibility is made salient for us. That possibility is that "I will rest well tonight and discover a cure for cancer in the morning": R & C. Since there is no chance of that happening but some chance we won't rest well tonight, we (fallaciously) conclude that (ii) is less likely than (i). The challenge for the rationalist is to provide evidence that nothing like this happens with (suitably

trained) philosophical intuitions: are we sure we reliably consider all the possibilities, not just the salient ones?

We can have no such assurance. On the rationalist's own view, we perforce consider only salient possibilities in the form of those we can conceive of. Are our concepts always adequate? It would be more appropriate to ask whether our concepts are ever adequate: the history of science reveals that fundamental concepts such as space, time, mass, living thing, and so on have undergone massive changes, such that the new concepts replacing the older ones cannot even be represented by the latter. If we can be certain of anything, we can be certain that our current concepts of mind, agency, rationality, consciousness, and other mental features will be forced to accommodate discoveries in cognitive neuroscience and will therefore have to change in significant and perhaps even radical ways.

The upshot is that even if our modal intuitions derive exclusively from our mastery of concepts, which naturalistic realists dispute, there is still no reason to believe that those intuitions are in any way reliable indicators of the modal facts: that a *prima facie* reason for believing that a certain state of affairs *S* is metaphysically possible (or impossible) is one's strong intuitive impression that *S* is possible (or impossible) or that a convincing demonstration that there is no *a priori* contradiction in supposing *S* obtains suffices to show that the world really could be in state *S*. Since our concepts are not transparent to us, we are not always reliable in computing the results of their interaction, including and especially the entailments they ground. Moreover, the concepts we use at any stage of inquiry suffer a natural attrition through the advance of science, becoming replaced by more comprehensive and refined successors. In short, if naturalistic realism is true, philosophical intuitions and the thought experiments that prime them are not reliable sources of *a priori* truths.

In conclusion, we have seen that realism and naturalism do indeed appear to be compatible with each other and that the combination of the metaphysics of realism with the epistemology of naturalism in naturalistic realism is an attractive philosophical position. However, naturalistic realism cannot be defended unless one is prepared to surrender some venerable assumptions about the nature of philosophy and the epistemological status of philosophical intuitions.

References

Button, T. (2013). *The Limits of Realism*. Oxford: Oxford University Press.

Brockman, J. (2011). The Marvels and the Flaws of Intuitive Thinking. In *Thinking*. New York: Harper Collins.

Carnap, R. (1963). Replies. In *The Philosophy of Rudolf Carnap*, edited by P. Schilpp. La Salle: Open Court.

Devitt, M. (1983). Realism and the Renegade Putnam: A Critical Study of Meaning and the Moral Sciences. *Noûs* 17(2): 291–301.

Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.

Dummett, M. (1991). *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.

Dummett, M. (1993). *The Seas of Language*. Oxford: Oxford University Press.

Dunning, D., Griffin, D.W., Milojkovic, J.D., and Ross, L. (1990). The Overconfidence Effect in Social Prediction. *Journal of Personality and Social Psychology* 57: 5–15.

Fischoff, B., Slovic, P., and Lichtenstein, S. (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perception and Performance* 3: 552–564.

Geach, P.T. (1966). Plato's "Euthyphro": An Analysis and Commentary. *The Monist* 50: 370–372.

Kahneman, D. (2012). *Thinking Fast and Slow*. New York: Penguin Books.

Khlentzos, D. (2004). *Naturalistic Realism and the Antirealist Challenge*. Cambridge, MA: Bradford Books/MIT Press.

Khlentzos, D. (2011). Challenges to Metaphysical Realism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/realism-sem-challenge/ (last accessed July 15, 2015).

Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61: 343–377.

Lewis, D. (1984). Putnam's Paradox. *Australasian Journal of Philosophy* 62: 221–236.

Lichtenstein, S. and Fischoff, B. (1977). Do Those Who Know More also Know More about How Much They Know? *Organizational Behaviour and Human Performance* 26: 149–171.

Lichtenstein, S., Fischoff, B., and Phillips, L.D. (1982). Calibration of Probabilities: The State of the Art to 1980. In *Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahnemann, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press.

Malmgren, A.S. (2011). Rationalism and the Content of Intuitive Judgements. *Mind* 120(478): 263–327.

Oskamp, S. (1965). Overconfidence in Case Study Judgments. *Journal of Consulting Psychology* 29: 261–265.

Putnam, H. (1985). *Philosophical Papers*, vol. 3: *Realism and Reason*. Cambridge: Cambridge University Press.

Putnam, H. (1992b). Reply to Gary Ebbs. *Philosophical Topics: The Philosophy of Hilary Putnam* 20: 347–358.

Putnam, H. (1994b). *Words and Life*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1982). *Methods of Logic*, 4th edn. Cambridge, MA: Harvard University Press.

Soames, S. (2006). The Philosophical Significance of the Kripkean *Necessary A Posteriori*. *Philosophical Topics* 16: 287–309.

Tversky, A. and Kahneman, D. (1982). Evidential Impact of Base Rates. In *Judgment under Uncertainty: Heuristics and Biases*, edited by D. Kahneman, P. Slovic, and A. Tversky, pp. 153–160. New York: Cambridge University Press.

Vallone, R.P., Griffin, D.W., Lin, S., and Ross, L. (1990). Overconfident Prediction of Future Actions and Outcomes by Self and Others. *Journal of Personality and Social Psychology* 58: 582–592.

12

Non-Naturalistic Metaphysics

HUD HUDSON

Metaphysics and Naturalism

There is no easy beginning here. Both "metaphysics" and "naturalism" are notoriously hard to define, and heading off in some direction illuminated by one manner of disambiguating our terms is sure to disappoint those who would have preferred exploring another path suggested by other available readings. To take up every combination, however, is to guarantee an overly general and unsophisticated treatment of our theme. What to do?

I will aim for a compromise. In the present section, I offer some general remarks on interpretations of the central terms. Then, in the two sections to follow, I will take up two rather different, specific, and intriguing ways one might interpret the topic on which I have something to say – in each instance critically discussing what role or hope there is for metaphysics (*in that sense*) against a backdrop of naturalism (*in that sense*). Along the way, I will offer some reflections about the resilience and even the inescapability of metaphysics, despite threats from naturalism to diminish its scope and significance.

What is it that all metaphysical problems or theses or inquiries or arguments have in common? Is it that their subject matter is confined to *being as such* or to *what there is* or to *first principles* or to *the supernatural* or to *the ultimate structure of reality* or to *fundamentality* or to *what grounds what*? Is it whatever Aristotle was up to in his *Metaphysics*, or what Kant was asking about the possibility of in his first *Critique*, or what Ayer boasted of eliminating in his *Language, Truth, and Logic*? Well, no. There are far too many paradigm cases of metaphysical problems, theses, inquiries, and arguments that would be misclassified as non-metaphysics should any of these candidates be proclaimed the single, unifying, core characteristic of that discipline.

Is metaphysics tied to a certain methodology, an approach largely characterized by a striving for precision, clarity, and consistency and aiming at properly formulating and evaluating the truth value of claims about a variety of subject matters by way of careful and sustained argument – or, as William James pithily put it, is metaphysics "nothing but an

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

unusually obstinate effort to think clearly" (James 1891, 145)? Again, no. That suggestion would misclassify as metaphysics much contemporary analytic philosophy in distinct fields such as ethics or epistemology or aesthetics, and it would misclassify as non-metaphysics many metaphysical explorations in nonanalytic traditions, and it would unreasonably disqualify a tremendous amount of metaphysics as practiced by many students on their midterm exams and by families at their dinner tables and by friends at their favorite coffee houses. Not all cases of engaging in metaphysics are instances of good philosophizing. Much of it falls so far short of the rosy description of methodology just described as to be truly awful. But awful metaphysics is metaphysics just the same.

I suspect that attempting to uncover necessary and sufficient conditions for being a bit of metaphysics is a challenge we are very likely to fail to answer satisfactorily. Better, I think, to identify and work with clear cases, asking about their connection with one or another form of naturalism, and letting the results trickle down to the borderline cases.

Initial inquiries into the meaning of "naturalism" are not much more promising. Clearly, naturalism is tied to a deep respect for science and for scientific method, to an equally deep suspicion of a variety of competing methods (especially if they smack of *a priori* theorizing about the world), and to something often approaching impatient contempt for religion and the supernatural (especially if it should intrude into the gaps currently left open by scientific theorizing about the world) – but *tied* in what way?

Is naturalism an ontological thesis about what does and does not exist – say, Armstrong's explicit insistence on naturalism as the view that the totality of what exists is nothing more than a unified, spacetime system, where the "nothing more" embodies a denial of transcendent gods, angels, demons, ghosts, souls, Cartesian minds, and all their supernatural ilk (Armstrong 1997, 5–6)?

Or is naturalism an epistemological thesis – say, Devitt's identification of naturalism with the austere view that the empirical way of knowing (i.e., the way that undergirds all science) is the only way of knowing (Devitt 1998, 45)? Of course, such a position could be conjoined with the ontological thesis just articulated, but it need not be; one could take naturalism to be a restriction on the sources and scope of knowledge without taking any stand whatsoever on whether those sources were adequate to investigate all the furniture of the world. Similarly, one could espouse a minimalist ontological thesis without endorsing the empirical-methods-only restriction on how we come to learn about our spacetime system and its inhabitants.

Or is naturalism a methodological thesis – say, Leiter's claim that theorizing in philosophy should be continuous with the empirical inquiry that characterizes the sciences (Leiter 1998, 81)? Such a methodology, of course, could be inspired by adherence to an ontological or epistemological version of naturalism, but its "should" could also simply be an expression of means–end reasoning, where some particular end regarding inquiry has been presupposed and advice is on offer about how to best achieve it. Again, the options may remain logically independent of one another.

All this to sort out – not to mention the hodgepodge of quasi-ontological or quasi-epistemological pronouncements about naturalism (i.e., those that lack any clear commitment to a substantive thesis about what the world contains or how we are to find out about it). Such claims are perhaps most charitably explained by taking naturalism to be a promising regulative principle, which instructs us to behave *as if* the metaphysical or epistemological theses just noted were true or *as if* the methodological thesis just noted were legitimate, while remaining silent about why.

Despite a similar proliferation of options, unlike in the case of "metaphysics," I believe there is a superior choice to be made when defining "naturalism." Rea has argued (convincingly, to my mind) that naturalism is best construed as a research program – something like a Quinean disposition (Quine 1995) to treat the methods of science as the only basic sources of evidence – on the grounds that the substantive ontological and epistemological theses that bear the name "naturalism" inevitably succumb to a dilemma that should appear fatal even by the lights of the proponents of those views (Rea 2002).[1] As such, naturalism would be opposed to other research programs identified with dispositions to recognize a narrower or wider range of basic sources of evidence, including, perhaps, rational intuitions or religious experience. A methodological naturalism, thus construed as a research program, can certainly be driven by significant sympathies with the ontological or epistemological forms of naturalism, as evidenced by Ladyman and Ross's polemical and influential recent book, which struck a nerve with its target audience and reignited a heated debate on these issues (Ladyman and Ross 2007; see also Ross, Ladyman, and Kincaid 2013).

In the end, however, I am not here seeking reasons to prefer one definition of "naturalism" over another; rather, I am exhibiting some popular ways of understanding that term in order to frame different discussions of non-naturalistic metaphysics.

We now have the materials to do so. In the next section, I will pair a methodological naturalism (construed as a research program and heavily inspired by the considerations associated with the epistemological naturalism sketched here) with the kind of work that is currently being practiced under the heading "contemporary analytic metaphysics." In the section following that, I will pair an ontological naturalism with the kind of work that could perhaps best be described under the heading "theistically informed metaphysics." Each pairing, then, gives us a window on the sort of confrontation to be had between a naturalism and a non-naturalized metaphysics.

The First Pairing: Methodological Naturalism and Contemporary Analytic Metaphysics

Exploring the tension in this first pairing leads some to the conclusion that there is simply no legitimate contribution for metaphysics to aspire to at all. The slightly less pessimistic may be led merely to the confidence that metaphysics should speak only when spoken to (by scientists), and then only to give modest advice about how to unify these or those bits of science. Less restrictive theorists may be inclined to a concession qualified by some cautionary advice, to wit, that whereas perhaps metaphysics can limp along and offer an interesting suggestion or two on its own initiative regarding questions underdetermined by science (maybe even chancing upon interesting lines of inquiry to be properly pursued with better tools), metaphysics should always remember to pay the strictest attention to science in order to keep from embarrassing itself, and in no case should it presume to pronounce on topics where science has already rendered a verdict, however tentative. And, of course, there are those who have come to see things the other way around, either acknowledging that metaphysics and science both provide serious and sophisticated approaches to learning

1 In sketching the alternatives that appear in these stage-setting remarks, I learned a great deal from Rea's (2002) excellent and penetrating discussion of the many varieties of naturalism and of the challenges that substantive versions of that thesis must confront.

about the world and that each voice should be heard, even when they conflict, or else insisting on a special autonomy and authority in metaphysics and tending toward the judgment that, whenever turf wars are at issue, metaphysics trumps science.

But how, exactly, does such a tension arise between methodological naturalism and contemporary metaphysics? As Humphreys has persuasively explained in a closely related piece on the similar tension between scientific ontology and speculative ontology (Humphreys 2013), the answer has a lot to do with (1) the increasingly popular belief that so much of what passes for metaphysical inquiry into and defense of allegedly live viewpoints is in fact vain and overexcited speculation about widely known falsehoods, with (2) the heavy reliance of contemporary metaphysics on intuitions and conceptual analysis (each suspicious on multiple grounds), and with (3) the careless presumption rampant among current practitioners of metaphysics of scale invariance that the physics of the very small is not radically different from what they take to be the physics of medium-sized, human-perceptible material objects. Especially irritating, apparently, is the habit of many metaphysicians of invoking scientific conceptions that are centuries old (and known to be riddled with problems even in their own day), despite having had their defects repeatedly explained to them and despite the easy accessibility of up-to-date science just a few Internet-clicks away.

It is, of course, possible to dispute cases. Some of the accused will retort that science has not conclusively shown, to take some recently contested examples, that there are no extended material objects or that there are no such things as spacetime points (or, if these theses have been shown, metaphysics was involved in the demonstration, as well). Additionally, they may note that up-to-date science is not an infallible oracle (as the *New York Times* reminds us every few months, when some significant breakthrough requires rewriting the textbooks in this or that field), thus hinting at a pessimistic induction as reason for their tardiness in keeping abreast of the current scientific journals. But such quibbles here and there should not lead anyone to dismiss this criticism of contemporary metaphysics too easily: science has a great deal to offer metaphysics, and to the extent that relevant scientific work in, say, cosmology, physics, or biology bears on some point at issue in metaphysical inquiries, the scientifically ignorant ignore it at their own peril.

It is more important, I think, to acknowledge that science has a good deal to teach metaphysics, but to recognize the difference between that clear truth and the somewhat less compelling thesis that metaphysicians always have a good deal to learn by listening to scientists. The difference, quite simply, is that scientists do not always speak *qua* scientists, but (as do all of us) occasionally venture out into waters that are decidedly metaphysical. In this modern age of science, we are inclined to closely attend to their opinions even in this context, accustomed as we are to deferring to their authority in areas where they can claim a level of expertise and a track record that are truly admirable. However, it is worth remembering that in these waters, a scientist often speaks *qua* metaphysician, and, more to the point, *qua* amateur metaphysician. There is no shame in that. There is no reason to think the expert in 16th-century poetry should be an authority in paleoanthropology, and there is no reason to expect the distinguished evolutionary biologist or theoretical physicist to be an expert in philosophy. Nevertheless, similar criticisms can go both ways: whereas the metaphysician is often scientifically illiterate, the scientist is often philosophically amateurish; whereas the metaphysician is frequently in danger of pursuing dead ends described in outdated scientific language, the scientist is frequently in danger of being unaware she has become a philosopher at all.

Notwithstanding this last point, certain prominent naturalists have become so frustrated with what they see as the unjustified pretensions of much of contemporary analytic metaphysics that two of their number have authored a book that opens by proclaiming that though this "professional activity [is] engaged in by some extremely intelligent and morally serious people, [it] fails to qualify as part of the enlightened pursuit of objective truth, and should be discontinued." The authors then proceed to canvas and ridicule the efforts of the scientifically uninformed, intuition-wielding metaphysicians of the hour who have been "wasting their talents [and] sowing systematic confusion about the nature of the world" (Ladyman and Ross 2007, vii). The proposed remedy? Naturalize metaphysics by seriously curtailing its excesses with the principle of naturalistic closure (PNC):

> Any new metaphysical claim that is to be taken seriously at time *t* should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than the sum of what is explained by the two hypotheses taken separately. (Ladyman and Ross 2007, 37)

The Ladyman and Ross proposal has generated a flurry of responses from analytic metaphysicians, as well as from those who sympathize with their frustrations. But even sympathizers have challenged the specific motivation for this unduly harsh plan for naturalizing metaphysics, together with its implication that metaphysics will thereby be unable to offer *any* measure of corrective pressure on the content of scientific claims, and they have also challenged the lines of reasoning for restricting acceptable metaphysics to the work of *unification* and for further restricting that unification to the domain of *science* alone (Melnyk 2013).

The metaphysician-targets of the remarks tend to think that Ladyman and Ross have overstated their claims regarding just which metaphysical topics have been rendered obsolete by current science and just which metaphysical theses enjoy overwhelming (or even majority) support among current metaphysicians, and especially about exactly what kind of work the appeal to intuitions is in fact designed to do in much of analytic metaphysics (Dorr 2010; Hawley 2010).

While the Ladyman and Ross proposal strikes me as an overreaction, it affords an important insight – namely, that philosophers would do well to increase their knowledge of science beyond A-level chemistry. However, this worthwhile advice does not justify gagging metaphysics until there is some science standing by in need of unification.[2] One modest kind of push back is to argue that metaphysics is not as different from science as it might appear at first blush. Both invoke similar forms of *a priori* and *a posteriori* reasoning, and whereas metaphysics and science have different subject matters – with the subject matter of metaphysics prior to that of science (as argued in Lowe 2002) – they nevertheless share a methodology (as defended in Paul 2012). On this optimistic conception, as Paul argues, metaphysicians need not be disparaged as armchair scientists, but neither need they be relegated to the role of handmaiden to science; there is perfectly respectable explanatory work aplenty to occupy metaphysics itself. Despite conceding considerable ground to science, however, resistance to even this friendly mode of reconciliation remains firmly in place (Ladyman 2012).

2 Perhaps some readers will remember the cartoon characters, Spike (an impressive bulldog) and Chester (his little, snappy companion), whose primary line seemed to be: "Whatta we gonna do today, Spike? Huh? Huh? Whatta we gonna do today?" PNC would appear to recast the cartoon with Science and Metaphysics.

Whether or not metaphysics and science can be reasonably thought to share a common methodology, metaphysics can claim a number of virtues and achievements of its own. Metaphysics is uniquely positioned to contribute to a number of debates that are historically of great interest to those who find methodological and epistemological naturalism attractive. Fruitful conceptual frameworks and exacting precision have been articulated and insisted upon by way of analytic metaphysics in the study of topics such as properties and relations, time, persistence, free will, causality, vagueness, mereology, category theory, grounding, modality, logic, and so on. Moreover, many of the standards of clarity that have been demanded and achieved in these discussions cannot fairly be laid at the feet of science. Finally, as is frequently remarked, metaphysics can contribute to the content of our theories about these topics in ways that are underdetermined by science (i.e., on those occasions when just which beliefs we should hold are not fixed by the scientific evidence at hand) (Sklar 1981; French 1998). In the philosophy of time, for example, presentists and growing block theorists are committed to a special notion of presentness that many believe has been put in jeopardy by the special theory of relativity, according to which there is no privileged reference frame suited to furnishing a special notion of presentness or legitimizing talk of absolute simultaneity; to the extent that these theories of time require one, so much the worse for them. This alleged threat is often rejected by such theorists of time on the grounds that whether or not appeal to a privileged reference frame would have any role to play in empirical science, it may have a perfectly good role to play in philosophical theorizing and in our attempt to explain the truth conditions of a variety of propositions we take to be true about time (Hawley 2006). Moreover, that role is role enough to lead us to consider seriously adding a certain kind of structure to our theories of time or even to consider proposing alternatives to the special theory of relativity on metaphysical grounds alone (Tooley 1997; Monton 2006).

Furthermore, metaphysics is often engaged in areas that don't compete with contemporary science. A fascinating puzzle about parthood and persons known as the "problem of the many," for example, is a wonderful little metaphysical case in point (for a full introduction to the puzzle, see Hudson 2001). It is often set up in a way that invites the charge of illegitimately assuming scale invariance, of treating macrophysical objects as fusions of particles and particles as tiny billiard balls smashing about in space (but, for the record, there is absolutely no need to frame the puzzle that way or to take any special view about how human persons or baseballs are constituted at the fundamental level, or even if there is such a thing as a fundamental level). It should come as an unnerving surprise to anyone (even to the ontological, epistemological, or methodological naturalist) that it is an open question how many people are in your chair right now, yet a count *much* higher than 1 is one of the negotiable resolutions of this problem! Escape from the puzzle by way of announcing that quantum physics has shown that there are no such things as bodies or baseballs to have puzzles about in the first place (Ladyman and Ross 2007, 301–302) is decidedly premature (Dorr 2010).[3] Moreover, the costs of such an escape will (or at least, should) come as *far more surprising news* than the puzzle it attempts to disarm.[4] And, finally, that this manner of escape is in the offing at all is quite unlikely to be the kind of

3 Or at least, this is how I anticipate Ladyman and Ross responding to the puzzle. Their purported acknowledgement of baseballs as "aspects of the world with sufficient cohesion at our scale that a group of cognitive systems with practically motivated interest in tracking them would sort them into types for book-keeping purposes" (2007, 5) isn't really an acknowledgement of baseballs.

4 Ladyman and Ross acknowledge this observation (2007, 302), but (one has to give them credit) regard swallowing this bitter pill as a sort of test to separate the real naturalists from the posers.

news that science-*sans*-metaphysics could discover on its own. It won't do to say that there is no fact of the matter, for the puzzle can be formulated as an inconsistent set of propositions, and any such shrug response will leave the inconsistency intact and invite questions about how the logic is to be adjusted. "From this perspective, non-naturalistic metaphysics is what happens when you take logic seriously."[5]

Similar puzzles regarding identity over time are especially resistant to responses of indifference or suggestions that the problems are not worth worrying about, yet they are underdetermined by scientific theories. Science will have volumes of information to offer on biological organisms and on psychological states and capacities, but ultimately it does not adjudicate between the claims of the biological and the psychological criteria for personal identity, and a fair amount of things the vast majority of naturalists care about hang in that particular balance (Hudson 2001, ch. 6). And, worth noting again, since many of the difficulties can be formulated as inconsistent sets of propositions, once these debates are acknowledged as meaningful, refusing to reject any of the propositions in question (for whatever reasons) calls for a revision in logic. But logic is the most general of the sciences, and naturalism must conform to logic (not the other way round). If we revise logical principles, the science that has been advanced on those principles will need to be revisited to determine whether we still know what we thought it taught us when those principles were doing their work unmolested and free of suspicion.

Of course, the naturalist may wish to respond either that the puzzles are not as easy to reconstruct in a scientifically respectable way as I am suggesting or that it is just missing the point to pretend that caring about a topic that science can't settle is a reason for thinking that contemporary analytic metaphysics is the right thing to throw at it. Nor, the naturalist may continue, *must* we fuss with logic to make palatable a dismissive attitude about the problems themselves. Perhaps we simply must admit our ignorance about some matters we happen to care about (unless science surprises us in days to come with unforeseen discoveries on the matter), or else come to realize that certain things we thought meaningful are really unintelligible after all.

In answering such a response from the naturalist, it is possible to hit closer to home and to raise the stakes of replying in this fashion. Contemporary science, in all its branches, is shot through with modality: claims about what is necessary, possible, or impossible. Contemporary metaphysics offers serious and sophisticated debates about modality concerning both what possibilities are (Lewis 1986; van Inwagen 1986) and which possibility-claims are true (a very common activity, despite the fact that such claims don't seem in any way to be empirically tractable – as argued in Rea 2002). It is much harder to feign ignorance about matters modal or to lump such claims in with putatively dubious worries about, say, free will or vagueness under the heading "bad or confused questions" without taking a noticeable hit in one's naturalist worldview as well.[6]

Given that the extrascientific theses aimed at in metaphysical theorizing are generally regarded as meaningful (i.e., the intelligibility of such claims is no longer as widely challenged as was fashionable a century ago), significant worries nevertheless remain about

5 A quote I borrow from Jason Turner, who provided very helpful suggestions in conversation on this section.

6 To be fair, Ladyman and Ross are aware of the cost of avoiding this criticism, and here too are ready to bite the bullet (2007, 300–302), but separating themselves from nearly all of their fellow naturalists by not trying to have it both ways and bravely following their argument to the utter abandonment of nontrivial modal properties can signal intellectual integrity and reductio at the same time.

cavalier appeals to intuitions and arguments to the best explanation (which themselves are often accused simply of being bare intuitive judgments masquerading as well-founded principles of theory selection). Dismissiveness of analytic metaphysics need not be grounded in the dubious complaint that its claims fail to be meaningful or in exaggerated boasts about what science has already conclusively shown us about, say, consciousness or causality. Rather, it can simply be the result of thinking that intuition-mongering armchair-philosophizing is incapable of filling in those (admittedly interesting) scientific blanks in reliable ways. Intuitions, we are reminded, are non-uniform among metaphysicians, display wildly different success rates in different domains, and are overly sensitive to environmental and historical factors (Humphreys 2013).

Champions of the intuition approach attempt to counter these reasons. One strategy is to note the difficulty in making explicit and then making stick the general charges of suspicion against the use of intuition. Such a task would involve identifying a normative premise (a necessary condition on justified belief), a non-normative premise (showing that appeal to intuition fails to satisfy this condition), a local constraint (explaining why the skepticism attaches to intuition alone, rather than to cognitive faculties much more generally), and a non-self-undermining constraint (ensuring that the justification of the premises just noted is not rooted in intuitions of the very sort that the argument seeks to undermine). Once the standards for a successful skeptical argument are in place, the defender of intuition can attempt to show that arguments from, say, a lack of independent calibration for intuition, from the non-uniform character of intuition, or from the unreliability of intuition fail to satisfy one or another of the relevant adequacy conditions on a successful skeptical argument (Pust 2012).[7]

The defender of intuition can also go on the offensive, arguing first that it is not at all clear that we can have a decent epistemology of logic without intuition (e.g., in rulings on just which argument forms are valid), and second that not only is intuition "the only source of justification for claims about justification, reason, evidence, and other epistemic concepts" (Pust 2012; a sentiment also defended in a series of articles by Bealer, e.g., 1998), but that it is, in particular, a self-supporting source of evidence.[8] Further, one can point to the fact that seemings are a genus with both perceptual and intellectual species, and, to the extent that we recognize a significant connection between seemings (at a general level) and *prima facie* justification, we may infer that intuitions have justificatory power just as do their better-respected cousins, perceptual seemings (Huemer 2005). In short, it is difficult to specially favor perceptual seemings while dismissing intellectual seemings (Chudnoff 2011), and justifying claims about epistemic concepts of any kind without straightforward reliance on intuitions of the very sort so often impugned by the methodological naturalist is an untenable, self-defeating task (Bealer 1992).

Methodological naturalism, as a disposition to treat the methods of science as the *only* basic sources of evidence, is either unacceptably impoverished or else misleadingly takes

7 The four conditions and their application to a host of specific arguments for intuition-skepticism are taken from Pust's (2012) excellent discussion of the nature of intuitions, of their epistemological role, and of arguments for and against their employment in different contexts of reasoning. I am also indebted to Pust for the brief overview of defenses of intuition that turn on considerations of epistemic circularity and similarities to perceptual justification given in the next paragraph.

8 Epistemic circularity threatens, but only in the guise that we have been taught to recognize and regard as innocuous inasmuch as it is the kind of epistemic circularity that attends any alleged foundational source of evidence (Alston 1993).

the phrase "the methods of science" to encompass somewhat more when actually invoked in practice than it does when announced as a corrective to the excesses of contemporary analytic metaphysics, put on a pedestal (where its foundations are harder to see), and admired. An ineliminable metaphysics whose subject matter is prior to that of science already lies at its base, and "provides the very framework within which such sciences are conceived and related to one another" (Lowe 2002). Substantive metaphysical theses inevitably make an appearance when discussions of properties, relations, time, causation, dependence, and modality are at issue, even if they are occasionally hard to recognize, dressed up in new categories and terminology (Hawley 2010). I side with Paul (2012) in thinking that the subject matter divides more than the methodology, but partly because I suspect that there is already a good deal more metaphysics in the practice and foundations of science (even in the form of reliance on the sorts of intuition whose credentials are forever being impugned) than is openly admitted.

The Second Pairing: Ontological Naturalism and Theistically Informed Metaphysics

Now suppose that naturalism is an ontological thesis about what does and does not exist. What, exactly, is in and what out? Though there is room for considerable debate on both sides, one common assumption is that God, at least, is out. Atheism certainly seemed to be intended in Armstrong's version of ontological naturalism, according to which the totality of what exists is nothing more than a unified, spacetime system. But what of the theist who interprets God's omnipresence and eternality as guaranteeing that God is literally present at each region of spacetime? Does such a conception of divinity then place God back among the potential candidates to be recognized by such naturalism? Oftentimes, the literature on these questions seems to suggest there is an implicit rider on such a conception of naturalism to the effect that "there is no good reason to believe in God or anything relevantly like God…in explaining or understanding the world" (Chapter 5). But, of course, one can entertain the hypothesis that God exists in the absence of good reasons for such belief (Clark 1990; Plantinga 2000). Since the common presumption appears to be that the door is closed on even this combination, however, perhaps the preferred rider should be taken to entail that whether conceived as transcendent, immanent, or even a bit of both, God simply does not exist, full stop.

Thus (popularly) construed, ontological naturalism quite clearly leaves no room whatsoever for theistically informed metaphysics.[9] So how should the theistic metaphysician (who accepts the boring but obvious truth that one should take into account whatever one believes exists when engaged in metaphysics) react to this ban on his livelihood? Well, one option is for the theist to convert to atheism and to start doing her work properly (i.e., without any reference to God's existence, nature, or activity). Another involves inquiring into the grounds of such an ontological naturalism and its preferred rider (whether implicit or not). Perhaps the pressure for atheistic conversion will diminish if inspection reveals naturalism to be an unstable (or even self-refuting) position by leading its proponents

9 I suppose it would still leave room for metaphysics informed by the fiction of theism; that is, for a strategy of investigating certain metaphysical problems and puzzles as if theism were true. Intriguing though it is, I set this consideration aside in what follows.

either to a conclusion of the global unreliability of our cognitive faculties (Plantinga 2000) or to the more modest (and more easily defended) conclusion of the unreliability of those of our cognitive faculties responsible for generating beliefs concerning "abstruse philosophical matters" like the truth or falsity of naturalism thus conceived (Rea 2002; Crisp 2011).

Another option for the theist is to concede that naturalism thus conceived is a stable position, but to inquire into and critically evaluate the grounds for its verdict of atheism. There are well-known, formidable, *philosophical* arguments for atheism (e.g., arguments rooted in claims about horrific and inscrutable evil or in claims about freedom and foreknowledge), and something of this sort may well be the justification for the atheistic rider on the ontological naturalism thesis. Interestingly, though, a defense along these lines would seem to run against the methodological and epistemological theses of naturalism at issue in the previous section, since philosophical arguments for atheism would require defense of a premise such as "there is no compensating good or morally justifying reason for the world's evils" or "we are sometimes free and morally responsible for our behavior despite the fact that divine foreknowledge is incompatible with these features"; that is, they would require defense of a premise that would not be demonstrable, given the restrictions of methodological naturalism.

In the remainder of this section, I will focus on a combination of views that, though exceedingly widely advocated, is, I believe, without much promise: namely, that ontological naturalism (combined with an explicit atheistic commitment) follows from epistemological naturalism together with the specific deliverances of modern science. On the contrary, agnosticism is a better companion for epistemological or methodological naturalism than is atheism, and in support of this pronouncement I will turn my attention briefly to a high-profile debate in the popular science/religion wars.[10]

The Darwinian take on evolution – descent with variation, together with the pressures of natural selection and chance – is nearly universally conceded to be the best explanation of the rich and intricate history of the tremendous complexity, diversity, and apparent purposiveness of the lifeforms that have been present on our planet for well over a billion years.

Darwin has made it possible, explains a grateful Dawkins, "to be an intellectually satisfied atheist" (Dawkins 1996). But how did Darwin accomplish such a thing? Even if an argument for theism from the best explanation for the origin of species (an argument that may have been rather compelling in the 18th century) was defeated by a competing Darwinian explanation, atheism doesn't follow from a bad argument for theism, and intellectual satisfaction for the atheist, one presumes, requires more than a naturalistic explanation of a set of observations about the history and diversity of lifeforms. There are, after all, a handful of other considerations to be accounted for.

One charitable reason that so many are inclined to approve of the "intellectually satisfied" sentiment is that it is heard not as the silly claim that "all the atheist's intellectual difficulties have their resolution in Darwinism," but rather as the promising claim that "the Darwinian take on evolution is inconsistent with God's existence, and thus, to the extent that one has adequate grounds for that scientific theory, one has intellectually satisfying grounds for atheism." Why *promising*? Because those in the know keep saying as much, and some deference to scientific expertise is certainly in order here.

I wholeheartedly agree: deference to scientific expertise on scientific matters is wise counsel. However, it may turn out that there are two (rather different) elements that are often combined

10 In the discussion that follows, I occasionally borrow from Chapter 3 of Hudson (2014).

under the heading "the Darwinian take on evolution" or "Darwinian evolutionary theory," and if we aim to defer to both on the basis of that wise counsel, we need to be assured that they are not a mix of science and speculative metaphysics (masquerading as science alone).

So, should the theist think that the Darwinian take on evolution is inconsistent with God's existence? Once this thesis is in play, what role could be left for God or for divine causality? For starters, here is a simple (but curiously overlooked) point: any theist who thinks that God is both omniscient and omnipotent, who thinks that the various stages of evolution are metaphysically contingent, and who thinks that refrainings are causes should straightforwardly believe that God is a causal contributing factor at every twist and turn of evolutionary development, from spectacular transitions to mundane mutations. This is not merely the familiar refrain that God always causally sustains creaturely things in existence from instant to instant, but rather the claim that insofar as God is able to interrupt or prevent any moment of the process and yet intentionally refrains from doing so, God is automatically among the causally contributing factors responsible for forging each and every link in the evolutionary chain.

More to the point, however, in addition to permitting evolutionary development to take its natural course, is God ever causally active in a more robust sense? Two questions are worth separating here: (1) Are there any good reasons to think that God cannot or does not ever precipitate a certain gene mutation, or steer a population toward a habitat in which it will especially thrive, or cause a predator to follow a false trail and lose its victim, or even invest an individual or a small group of individuals with powers and capacities that that have no natural, antecedent causes? (2) Are there any good reasons to think that God can or ever does do such things?

All too often, (1) is answered affirmatively on the grounds that scientific progress itself has decisively ruled out any such divine participation (Dawkins 1996; Dennett 1996; Kitcher 2009). Allegedly, then, this "discovery" can serve as the source of a new and noncircular argument for atheism.

Yet very *little* is actually presupposed in resisting this line of reasoning. No one need contest that the earth is ancient and has hosted life for well over a billion years, nor that most of the astonishing variety of organisms that have been in existence at one time or another in that interval are now extinct, nor that there is a single tree of life such that any two living things can trace their lineage back to a common ancestor in the remote past, nor that the primary natural mechanism of the evolutionary process is variation in descent with natural selection. Whereas many theists have contested the single-tree-of-life thesis with a doctrine of "novelty creation" or have repeatedly raised doubts about whether modification and natural selection are the only primary (major) natural mechanisms in play (van Inwagen 1994; 1995), we can simply spot evolutionary biology *all* of the discoveries just noted. What cannot be simply granted (and what obviously should not be simply granted) is an additional, nonscientific, metaphysical thesis: namely, that God did not play any role in guiding the course of evolution by, say, determining which variations would occur in some given population. The additional claim is not the sort of thesis that can be established by science alone. Perhaps science has already established (or, for the skeptical, will be able to establish) that the natural mechanisms in question *could* produce the unity and variety of life the world has known without any divine intervention, but that is a far cry from showing that they actually *did* so.

The point is a significant and subtle one. Kitcher, for example, characterizes an objection that he calls "antiselectionism" as "the least ambitious of the challenges to Darwinism," and

explains that it consists in maintaining that natural selection doesn't have the *power* to do the work it is assigned, that it *couldn't* achieve those ends. He then spends a substantial chapter arguing that the contested-ability claim is true after all, even if significantly improbable. He concludes that even "the least ambitious challenge" can be put to rest (Kitcher 2009, ch. 4). But the theist can happily concede that further point as well; it need not be an issue. The point is not to question whether it *could* have happened within the timeframe allotted (absent divine direction and given the natural mechanisms proposed), but whether it *did* happen that way. And there is little to be said (short of independent philosophical arguments for atheism, which are being kept at bay in deference to the usual epistemic restrictions of the methodological naturalist) for moving from the premise that it could have happened in that terribly improbable way to the conclusion that it did. If it did, science alone is not in a position to say so (Plantinga 2011).

On the other hand, it *should* be commonplace (among theists) to answer (2) in the affirmative. Theistically guided evolution should appear (to the theist) overwhelmingly attractive as a leading candidate for God's mechanism for producing life, its unity, and its diversity (Plantinga 1993; Nagel 1997; Collins 2003; Clark 2014). (That's not to say anything on why it would be a defensible mechanism rather than an unnecessarily long, pointless, and cruel one; but, of course, it wasn't meant to. The moral justifiability of such a means of production belongs to another part of the comprehensive response to the problem of evil.) The most frequently encountered objections to the standard evolutionary account are alleged worries about irreducible complexity, insufficient timeframes for major transitions, and the inadequate evidential base of the fossil record. Let us assume that we may set aside objections from the fossil record on the grounds that the conditions for fossilization are rare enough that, whereas the theory may predict vast numbers of intermediate forms, it need not predict that we should have happened upon considerably more intermediate forms than we in fact have. The complexity-based and timeframe problems continue to surface in one form or another, however. The scientific consensus seems to be that whereas there are, indeed, some outstanding difficulties of each kind, they are answerable. Wrestling with whether such confidence is justified has spawned a sprawling literature of replies and counter-replies, but our present point only requires establishing one smallish side note in that acrimonious debate.

Suppose one is both *already* a theist (and thus grants the existence of an omnipotent and perfectly good being) and has also conceded the claim that the natural mechanisms on offer *could* accomplish the tasks in question, despite the considerable improbability of it all. Why would that *theist* be hesitant to believe that God was involved in directing the process? It's a bit like discovering the cards on your desk arranged in alphabetical suit, ace-to-king, recognizing that they could have come to occupy that pattern by random shuffling or by way of the intentional agency of your officemate (whom you believe had the power and desire and opportunity to so arrange them), and then puzzling over which option is the more reasonable hypothesis of the two.

Misunderstanding is a hazard at this point. *To be clear*: I am not arguing that any outstanding or widely discussed difficulty for evolutionary theory should make one throw one's hands in the air and turn to God as the only solution. Far from it. On the contrary, I'm conceding to evolutionary theory everything save the gratuitous, metaphysical addition that God doesn't ever participate in the process. Moreover, I'm suggesting that any theist who is already willing to concede these points should also be inclined to think evolution is divinely guided on the grounds that this combination view answers at no additional cost (to one who is *already* a theist) a number of otherwise perplexing questions.

The primary point of these reflections is to emphasize the thought that once epistemological or methodological naturalism is taken as fixed, "an atheistic worldview" is hardly an appropriate subtitle for the body of work that can emerge from that restricted base. Atheistic arguments rely on non-empirically testable premises and intuition judgments just as much as theistic ones do. Without question, science has generated serious obstacles for a host of theistic arguments, but demonstrating that some theistic argument is unsound doesn't license the assertion of the negation of its conclusion. Any ontological naturalism that carries the rider that God does not exist threatens theistically inspired metaphysics, to be sure, but it is also an ontological naturalism supported by an evidential base considerably richer than that of epistemological or methodological naturalism. The theorist interested in engaging in theistically informed metaphysics can then take heart, for the same rich evidential base allegedly justifying ontological naturalism is certain to contain the resources for combating it as well.

Acknowledgments

For criticisms and comments on earlier drafts of this paper, I thank Ross Cameron, Kelly James Clark, Daniel Howard-Snyder, Mike Rea, Jason Turner, Ryan Wasserman, and Dennis Whitcomb.

References

Alston, W. (1993). *The Reliability of Sense Perception*. Ithaca: Cornell University Press.

Armstrong, D. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.

Bealer, G. (1992). The Incoherence of Empiricism. *Proceedings of the Aristotelian Society, Supplementary Volume* 66: 99–138.

Bealer, G. (1998). Intuition and the Autonomy of Philosophy. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, edited by Michael Depaul and William Ramsey, pp. 201–240. Lantham: Rowman and Littlefield.

Chudnoff, E. (2011). The Nature of Intuitive Justification. *Philosophical Studies* 153: 313–333.

Clark, K.J. (1990). *Return to Reason: A Critique of Enlightenment Evidentialism and a Defense of Reason and Belief in God*. Grand Rapids: Eerdmans.

Clark, K.J. (2014). *Religion and the Sciences of Origins*. New York: Palgrave Macmillan.

Collins, R. (2003). Evolution and Original Sin. In *Perspectives on an Evolving Creation*, edited by Keith B. Miller, pp. 469–501. Grand Rapids: Eerdmans.

Crisp, T. (2011). An Evolutionary Objection to the Argument from Evil. In *Evidence and Religious Belief*, edited by Kelly James Clark and Raymond Van Arragon, pp. 114–134. Oxford: Oxford University Press.

Dawkins, R. (1996). *The Blind Watchmaker*. London: W.W. Norton.

Dennett, D. (1996). *Darwin's Dangerous Idea*. New York: Touchstone.

Devitt, M. (1998). Naturalism and the A Priori. *Philosophical Studies* 92: 45–65.

Dorr, C. (2010). Review: *Every Thing Must Go: Metaphysics Naturalized*. *Notre Dame Philosophical Reviews*. Available from: http://ndpr.nd.edu/news/24377-every-thing-must-go-metaphysics-naturalized (last accessed July 15, 2015).

French, S. (1998). On the Withering Away of Physical Objects. In *Interpreting Bodies*, edited by Elena Castellani, pp. 93–113. Princeton: Princeton University Press.

Hawley, K. (2006). Science as a Guide to Metaphysics. *Synthese* 149: 451–470.

Hawley, K. (2010). Throwing the Baby Out with the Bathwater. *Metascience* 19.2: 174–179.
Hudson, H. (2001). *A Materialist Metaphysics of the Human Person*. Ithaca: Cornell University Press.
Hudson, H. (2014). *The Fall and Hypertime*. Oxford: Oxford University Press.
Huemer, M. (2005). *Moral Intuitionism*. New York: Palgrave Macmillan.
Humphreys, P. (2013). Scientific Ontology and Speculative Ontology. In *Scientific Metaphysics*, edited by Don Ross, James Ladyman, and Harold Kincaid, pp. 51–78. Oxford: Oxford University Press.
James, W. (1891). *The Principles of Psychology*, vol. 1. London: Macmillan.
Kitcher, P. (2009). *Living with Darwin: Evolution, Design, and the Future of Faith*. Oxford: Oxford University Press.
Ladyman, J. (2012). Science, Metaphysics, and Method. *Philosophical Studies* 160: 31–51.
Ladyman, J. and Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
Leiter, B. (1998). Naturalism and Naturalized Jurisprudence. In *Analyzing Law: New Essays in Legal Theory*, edited by Brian Bix, pp. 79–104. Oxford: Clarendon Press.
Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.
Lowe, E.J. (2002). *A Survey of Metaphysics*. Oxford: Oxford University Press.
Melnyk, A. (2013). Can Metaphysics Be Naturalized? And If So, How? In *Scientific Metaphysics*, edited by Don Ross, James Ladyman, and Harold Kincaid, pp. 79–95. Oxford: Oxford University Press.
Monton, B. (2006). Presentism and Quantum Gravity. In *The Ontology of Spacetime*, edited by Dennis Dieks, pp. 263–280. Amsterdam: Elsevier.
Nagel, T. (1997). *The Last Word*. Oxford: Oxford University Press.
Paul, L.A. (2012). Metaphysics as Modeling: The Handmaiden's Tale. *Philosophical Studies* 160: 1–29.
Plantinga, A. (1993). *Warrant and Proper Function*. Oxford: Oxford University Press.
Plantinga, A. (2000). *Warranted Christian Belief*. Oxford: Oxford University Press.
Plantinga, A. (2011). *Where the Conflict Really Lies: Science, Religion, and Naturalism*. Oxford: Oxford University Press.
Pust, J. (2012). Intuition. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/win2012/entries/intuition/ (last accessed July 15, 2015).
Quine, W.V.O. (1995). Naturalism; or, Living within One's Means. *Dialectica* 49: 251–262.
Rea, M.C. (2002). *World Without Design: The Ontological Consequences of Naturalism*. Oxford: Oxford University Press.
Ross, D., Ladyman, J., and Kincaid, H. (eds.) (2013). *Scientific Metaphysics*. Oxford: Oxford University Press.
Sklar, L. (1981). Time, Reality, and Relativity. In *Reduction, Time, and Reality*, edited by Richard Healey, pp. 129–142. Cambridge: Cambridge University Press.
Tooley, M. (1997). *Time, Tense, and Causation*. Oxford: Oxford University Press.
van Inwagen, P. (1986). Two Concepts of Possible Worlds. *Midwest Studies in Philosophy* 11: 185–213.
van Inwagen, P. (1994). Doubts about Darwinism. In *Darwinism: Science or Philosophy?*, edited by Jon Buell and Virginia Hearn, pp. 177–191. Richardson: Foundation for Thought and Ethics.
van Inwagen, P. (1995). Genesis and Evolution. In *God, Knowledge, and Mystery*, pp. 128–162. Ithaca: Cornell University Press.

13

Naturalism and Physicalism

BARBARA GAIL MONTERO AND
DAVID PAPINEAU

Introduction

Naturalism means different things to different people. But one significant strand in contemporary understandings of the term is *physicalism*. This is the doctrine that everything is physical. In this chapter, we shall examine this doctrine and assess the strength of the case in its favor.

Physicalism has increased markedly in popularity in the Western world over the past century or so. In a recent survey of philosophers, 56% of the 3000-plus respondents were in favor of physicalism, and only 27% definitely against.[1] This is a relatively new phenomenon. The proportions would have been very different in the 19th century. One issue to be addressed in this chapter is the explanation of this shift.[2]

As understood by contemporary philosophers, physicalism is a relatively laissez-faire doctrine. The basic thought is that everything is physically constituted. But this is generally understood in such a way as to avoid any strong methodological implications. Few contemporary physicalists would argue that the truth of physicalism means that we should close down the psychology, biology, or even meteorology departments and hand everything over to the physicists. The practicalities of studying complex structures like minds, bodies, and weather systems call for special methods and techniques beyond those used in the analysis of basic physical processes. Moreover, many physicalists would add that there is a metaphysical basis for this methodological precept. In their view, the properties and patterns that are displayed in the "special sciences" (psychology, biology, meteorology, etc.) are genuinely novel, in that they cannot be defined within the language of basic physics or explained by basic physical principles.

1 http://philpapers.org/surveys/results.pl (last accessed July 15, 2015).

2 For a detailed account of how attitudes toward physicalism in the Western world have changed over the past four centuries, see Papineau (2001).

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

We shall return to some of these niceties further on. For the moment, it will suffice to characterize physicalism graphically. Imagine God creating the world, and in particular imagine that God's first task is to put all the physical material in place. Now we can ask: Given that God has arranged all the quarks, leptons, and so on, is it time to rest? Antiphysicalists will say "no," but physicalists will answer "yes." The antiphysicalists will feel that God still needs to add all the conscious thoughts and feelings. But physicalists will think that this has already been taken care of. By fixing the basic physical facts, God has therewith fixed all the facts, including conscious mental facts. Nothing more is needed for conscious minds than the relevant physical facts. Or, to put it the other way round, not even God could create a world physically just like ours but lacking conscious minds.[3]

This metaphor captures the technical notion of physicalism as the "metaphysically necessary supervenience" of all facts on the basic physical facts. This is the idea that everything is fixed once the fundamental physical facts are fixed: people and plants (and other such things), even though they are not necessarily reducible to quarks and leptons (and other such things), are in some sense nothing over and above quarks and leptons (and other such things). We shall understand "physicalism" in this sense in what follows. In line with this, we shall count some property or entity as "physical" not only when it is a basic physical property or entity, but also when it supervenes on the basic physical facts.

The Case for Physicalism

Despite the previously stated qualifications, physicalism is still a very strong doctrine, and indeed one that has been denied throughout much of history. Why do so many contemporary philosophers embrace it? As we see it, the driving motivation behind the commitment to physicalism is the need to explain how things that are apparently not physical can have physical effects. Thus, many contemporary thinkers adopt a physicalist view of the mental realm because they think that otherwise we would be unable to explain how mental processes can causally influence the physical world. Similar considerations motivate physicalist views of the biological and other realms.

It may not be immediately obvious why this need to account for physical influence should demand that we view the mental, biological, and other realms as themselves physical. After all, there seems nothing *a priori* incoherent in the idea of nonphysical agents exerting a causal influence on physical processes, as is testified by the conceptual cogency of traditional stories in which active spirits and other immaterial agents intervene in the physical world.

However, there may be *a posteriori* objections to such nonphysical interventions, even if there are no *a priori* objections. We shall see further on how modern scientific theory places strong restrictions on the kinds of entities that can have physical effects. Given that mental and biological phenomena clearly do have such effects, this suggests that they must satisfy the relevant restrictions.

We can put the argument like this. Science shows us that physical effects can always be accounted for by fully *physical* causes. But we know that biological and mental facts are

3 This metaphor is inspired by Kripke's discussion of the mind–brain identity theory in *Naming and Necessity* (1980, 153–154). Though most accept this account of physicalism, see Montero (2013) for an argument that physicalism is consistent with the denial of such a view.

often among the causes of physical effects (as when a dog's *breathing* reduces *the oxygen level,* or when I *decide to move that stone* and thereby *move it*). So those biological and mental causes must themselves be fully physical.

In effect, the scientific discovery that physical effects always have physical causes squeezes any *non*physical factors out of the realm of things that can affect the physical world. So, in order to account for how mental and biological processes do affect the physical world, we have to recognize that they are themselves physical.

Physicalists do not deny that there are mental and biological facts. Of course there are. They don't want to eliminate these facts, but illuminate their nature. In this spirit, they view mental and biological facts as physical. When we talk about *decisions* or *breathing*, we aren't talking about some extra facts, distinct from electrochemical goings-on in brains and bodies. Rather, we are just talking about large-scale aspects of the underlying electrochemical processes.

Note how the science-based argument outlined in this section (if mental and biological causes weren't physical, they would be squeezed out of the range of things that can affect the physical world) is different from some of the more traditional arguments against Cartesian dualism and similar nonphysicalist philosophies. Some of the earliest commentators on Descartes argued that he had divided mind and body too sharply to allow any causal interaction between them. It is not clear how telling this worry is. On many conceptions of causation there is no reason why there should not be causal intercourse between Descartes' two realms, and historically, as we shall see, it seems unlikely that this traditional concern did much to discredit Cartesian interactionist dualism. In any case, the science-based argument on which we shall focus is different. Its thrust is not that immaterial minds and other nonphysical entities are *a priori* the wrong kind of thing to affect the physical world, but simply that science has shown us *a posteriori* that the physical world is not subject to such influences.

Note also how the science-based argument only indicates that the types of things that have physical effects are themselves also physical; it says nothing about types of things that have no physical effects. For example, many philosophers think that there are abstract numbers and sets which inhabit some realm outside space and time and so are causally inert and lack any physical effects. Our science-based argument leaves it open whether such things must be physical. Accordingly, we shall understand "physicalism" in what follows not as the thesis that absolutely *everything* is physical (as we put it at the start), but as the more qualified claim that everything within the spatial or temporal realm is physical.

Leibniz, Newton, and the Conservation of Energy

It will be worth explaining in some detail the evolution of modern scientific ideas about the range of things that can have physical effects. This will help to forestall the impression that contemporary physicalism is some kind of fad. Thus, it is sometimes suggested that physicalism rests, not on reasoned argument, but on some kind of unargued commitment, some ultimate decision to nail one's colors to the physicalist mast. And this diagnosis can seem to be supported by the fact that physicalist doctrines have become widely popular in the Western world only in the past few decades. However, familiarity with the relevant scientific history casts the matter in a different light. It turns out that physicalist doctrines,

far from varying with ephemeral fashion, are closely responsive to received scientific opinion about the range of causes that can have physical effects.

Let us begin with the "mechanical philosophers" of the 17th century, who held that any material body maintains a constant velocity unless acted on, and moreover that all action is due to the impact between one material particle and another. So stated, the mechanical philosophy immediately precludes anything except impacting material particles from producing physical effects. Leibniz saw this clearly, and concluded that it discredited Descartes' interactive dualism, which had a nonmaterial mind influencing the physical world (Woolhouse 1985). (As it happens, Leibniz did not therewith reject dualism, but instead opted for "preestablished harmony." Views which avoid physicalist views of the mind by denying its causal efficacy will be discussed further on.)

At the end of the 17th century, Newtonian physics replaced the mechanical philosophy of Descartes and Leibniz. This reinstated the possibility of interactive dualism, since it allowed that disembodied nonimpact forces could cause physical effects. Newtonian physics was quite open-ended about the kinds of forces that exist. Early Newtonians posited distinctive mental and vital forces alongside magnetic, chemical, gravitational, and impact forces. Accordingly, they took fundamental mental action and fundamental vital action in the material world to be perfectly consistent with the principles of physics. Moreover, there is nothing in the original principles of Newtonian mechanics to stop fundamental mental forces arising autonomously and unpredictably, in line with common assumptions about the operation of the mind (Papineau 2001). As a result, the Newtonian world view was effectively an interactive pluralism that recognized a wide range of nonphysical influences, including spontaneous mental influences (or "determinations of the soul," as they would then have been called).

In the middle of the 19th century, the conservation of kinetic plus potential energy came to be accepted as a basic principle of physics (Elkana 1974). In itself, this does not rule out fundamental mental or vital forces, for there is no reason why such forces should not themselves be "conservative," operating in such a way as to compensate losses of kinetic energy by gains in potential energy and vice versa.[4] (The term "nervous energy" is a relic of the widespread late 19th-century assumption that mental processes store up a species of potential energy, which action then converts into the kinetic energy of bodily moments.) The conservation of energy, however, implied that such fundamental special forces must be governed by strict deterministic laws: if mental or vital forces arose spontaneously, then there would be nothing to ensure that they never led to energy increases. This had a great impact on 19th-century thought about free will. The idea that all mental and vital processes must be entirely governed by deterministic laws was viewed by many as incompatible with the traditional view of free agents as autonomous influences operating independently of the constraints of natural law.

It is an interesting question whether the 19th-century view that all fundamental mental or vital forces must be governed by deterministic laws already amounts to a doctrine worth calling "physicalism." In favor of this way of viewing things, the doctrine does portray all causally significant properties, including mental and vital ones, as within the realm of scientific theory and in principle subject to the kind of mathematical treatment familiar from the analysis of gravity and electromagnetism. But on the other side, the mental and

4 For discussion of the compatibility of fundamental mental forces with the conservation of energy law, see Montero (2006).

vital forces that it allows are fundamental causal agents, found only in sentient and living organisms, and additional to any forces operating in the inanimate world. Some will feel a doctrine that countenances such fundamental mental and vital forces is not worth counting as "physicalist."

Strong Physicalism Vindicated

As it happens, there is no great need to determine whether being governed by mathematically formulable deterministic laws suffices for an entity to count as physical. This is because 20th-century science has given us reason to suppose that there are no fundamental vital or mental forces after all, and that the only things capable of producing physical effects are entities constituted by the kind of basic force fields that can be found throughout the inanimate world (such as gravity, electromagnetism, and nuclear forces). Such science thus supports the stronger view that everything is made of entities that are not only governed by deterministic mathematical laws, but are also all found in inanimate realms. In other words, it supports the view that the underlying nature of humans and other creatures with minds in not different in kind from the underlying nature of inanimate things such as rocks. (In line with this, and given that the term "physical," when used in debates over physicalism, is a philosophical term of art, we will now stipulate that physical entities are those found in inanimate realms and those composed out of such entities.)

The argument against fundamental vital and mental forces is a simple empirical one. By the 1950s, it had become difficult to continue to uphold the existence of special vital or mental forces: detailed physiological research, especially into nerve cells, gave no indication of any physical effects that cannot be explained in terms of the basic physical forces that also occur outside living bodies. A great deal became known about biochemical and neurophysiological processes, especially at the level of the cell, and none of it gave any evidence for the existence of special forces not found elsewhere in nature.

Thus, during the first half of the century, the catalytic role and protein constitution of enzymes were recognized, basic biochemical cycles were identified, and the structure of proteins was analyzed, culminating in the discovery of DNA. In the same period, neurophysiological research mapped the body's neuronal network and analyzed the electrochemical mechanisms responsible for neuronal activity. Together, these developments made it difficult to go on maintaining that special forces operate inside living bodies. If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of any nonphysical forces. The underlying nature of human beings, this research indicated, is no different from the underlying nature of ordinary nonliving things (Papineau 2001).

This argument rests on normal inductive grounds. There is no principled *a priori* reason why 20th-century physiological research should not have uncovered special mental and vital forces. It is just that the inductive evidence went the other way. Of course, it is possible to resist the conclusion. You could continue to believe that there are special vital mental forces that operate in as yet undetected ways in the interstices of living tissues and intelligent brains, and resist physicalism on those grounds. But there seems little merit in this position. The nonexistence of special vital and mental forces has been established by over a century of detailed empirical research. Given this, it seems more rational to explore the consequences of this finding, rather than resist it.

Hempel's Dilemma

Some philosophers seek to cast doubt on physicalism by appealing to "Hempel's dilemma." In the 1950s, Carl Hempel argued that physicalism falls at the first hurdle on the grounds that there is no good way of understanding the term "physical." We might take "physical" to refer to just those entities that are recognized by contemporary physics. But then the doctrine that *everything is physical* will certainly be false, for we can be confident that contemporary physics is by no means the last word about the basic constituents of reality. Alternatively, we might understand "physical" as referring to just those entities recognized in the future by the successors of our contemporary theories, at the ideal end of enquiry, perhaps. But then the doctrine that *everything is physical* will be empty, for we have little idea what future physics will reveal.[5]

However, notwithstanding Hempel's suggestion, there is no need to define "physical" in terms of physical theory, either contemporary or future, to pick out some privileged notion of "physical." In truth, there are a number of *alternative* ways of defining "physical", *all* of which give rise to interesting doctrines arguably worth calling "physicalism." All that is needed is *some* way of identifying a category of facts (call them "Q") that satisfies the following requirements: (1) at first sight mental, biological, and similar categories do not seem to be Q, but (2) mental, biological, and similar facts nevertheless do have effects among Q-facts, while at the same time (3) there is good reason to think that Q is "causally complete": that is, that Q-effects always have fully Q-causes. As soon as we have a Q-category that satisfies these specifications, we can illuminatingly argue as before that mental, biological, and similar categories must after all be composed of Q-facts, despite first appearances – for otherwise how could they cause their Q-effects, given the completeness thesis that Q-effects always have fully Q-causes?[6]

As it happens, we have already identified *two* different categories satisfying these specifications. We explained earlier how the 19th-century discovery of the conservation of energy showed that the category of facts "falling under deterministic mathematically formulable laws" satisfies a completeness requirement, in that the discovery implied that all effects of this kind must have similar causes (in the form of conservative force fields). The implication was thus that mental and other such facts must themselves fall under deterministic mathematical laws, despite initial appearances to the contrary. And then we pointed out that 20th-century physiology gives us strong reason to think that the realm of phenomena composed of the kinds of entities found in inanimate realms is also causally complete, in that even within brains and bodies effects always seem to be produced by just the same kinds of electrochemical and other causes as operate in the inanimate realm. And this then argued that mental and other features of the world that have inanimate effects, in addition to falling under deterministic mathematical laws, must also be composed of the kinds of entities found in the inanimate realm.

Other philosophers run this form of argument with yet other specifications of a Q-category, such as "determined by microscopic components" or, again, "similar to the kind of entities recognized by current physics." To the extent that these categories satisfy the relevant specifications, and in particular a completeness requirement that effects of these kinds always have full causes of these kinds, they allow further interesting conclusions.

5 For discussion of Hempel's dilemma, see Montero (1999).

6 See Papineau and Spurrett (1999).

Do all of these conclusions line up with views that have traditionally been thought of as forms of physicalism? Some do, some might not, but, as we said earlier, "physicalism" is a term of philosophical art that philosophers use in a variety of ways. Our point is merely that any category that fits our criteria can be plugged into the causal argument to produce surprising results. Hempel's dilemma is avoided because the role of science, including physical science, is not, on our account, to give a definition of "physical," but rather to tell us whether the so-defined "physical" realm is causally complete. And it turns out that science supports the completeness claim for a range of differently defined "physical" realms.

It is important to realize that science does not have to tell us everything in order to tell us anything. Science has not yet, of course, verified a complete definitive list of the fundamental entities responsible for effects in the natural world. But this does not mean that it has not yet verified any significant facts about that list. And our historical analysis indicates that it has indeed established a significant amount of such information; for example, that those fundamental agents are all governed by deterministic mathematical laws and all operate inter alia in the inanimate realm (and plausibly also that they are all determined by microscopic components and all similar to the kinds of entities recognized by current physics).

So we have a number of alternative possible "physicalisms" to choose among, all of which give us highly interesting conclusions about the constitution of the world, and in particular about the constitution of mental and biological agents. In the interests of clarity and brevity, however, let us just stick to the definition we settled on earlier, and continue to understand "physical" as referring specifically to items composed of entities that are also found within the inanimate world.

The Rise of Physicalism in the 1950s and 60s

It is striking how physicalism emerged suddenly as a popular philosophical doctrine in the middle decades of the 20th century. From the 1950s on, a number of philosophers put forward arguments designed to establish physicalist conclusions, and in particular to show that the mind must be identified with the brain. Not all these arguments may seem to fit the analysis of physicalism we have given so far. While some of them did appeal explicitly to the causal closure of the physical realm (Feigl 1958; Oppenheim and Putnam 1958), other arguments from this period seemed to proceed rather differently.

However, it is not difficult to show that appearances are deceptive here. On examination, it turns out that even those arguments that don't explicitly appeal to causal closure do so implicitly, in line with our thesis that it was the empirical evidence for causal closure that persuaded philosophers to be physicalists. Once mid-century physiological research had established that all physical effects had physical causes, even in bodies and brains, philosophers quickly figured out that general physicalism followed (Papineau 2001).

Thus, for example, consider Smart's (1959) thought that we should identify mental states with brain states, for otherwise those mental states would be "nomological danglers" that play no role in the explanation of behavior. Similarly, reflect on Lewis (1966) and Armstrong's (1968) arguments that, since mental states are picked out by their causal roles, including their roles as causes of behavior, and since we know that physical states play these roles, mental states must be identical with those physical states. Or, again, consider Davidson's (1970)

argument that, since the only laws governing behavior are those connecting behavior with physical antecedents, mental events can only be causes of behavior if they are identical with those physical antecedents. There is much to say about these arguments. But the point we want to make here is that none of them is even slightly plausible without the assumption of the causal closure of physics.

To see this, imagine that the causal closure of physics were not true, and that some physical effects (the movements of arms, perhaps, or the firings of the motor neurons that instigate those movements) were not determined by law by prior physical causes at all, but instead by fundamental dualist mental causes, such as decisions or exercises of will, or perhaps just pains. If this were the case, then (1) contra Smart, mental states wouldn't be "nomological danglers," but would be directly efficacious in the production of behavior; (2) contra Lewis and Armstrong, it wouldn't necessarily be physical states that played the causal roles by which we pick out mental states, but quite possibly the *sui generis* mental states themselves; and (3) contra Davidson, it wouldn't be true that the only laws governing behavior are those connecting behavior with physical antecedents, since there would also be laws connecting behavior with mental antecedents.

Arguments against Physicalism

The contemporary philosophical literature contains a number of arguments against physicalism about the mental realm, and in particular about *conscious* mental states. Should they make us reconsider our stance on physicalism?

Central to the antiphysicalist literature is Frank Jackson's "knowledge argument" (1982), which takes up themes from Thomas Nagel's influential earlier article "What Is it Like to Be a Bat?" (1974). These authors point out that someone could know all there is to know from a third-person scientific point of view about certain conscious states yet not "know what it is like" to undergo them. Thus, Nagel observed that even an expert bat scientist could not know what it is like to echolocate, and Jackson constructed a thought experiment in which a fully scientifically informed but color experience-deprived vision scientist could not know what it is like to see colors.

Whether these examples discredit physicalism hinges on what is involved in "knowing what it is like" to undergo some experience. If this required knowing about some nonphysical feature of the world, then physicalism would certainly be false.

But it is contentious whether talk of "knowing what it is like" carries this implication. Nearly all contemporary physicalists allow that someone who has never undergone a certain kind of experience is cognitively limited as a result, and that a color experience-deprived vision scientist would learn something knew when she saw color for the first time. But they do not concede that this change involves her coming to know about something nonphysical. There are many subtle differences in how physicalists try to explain the change in the scientist. However, they all agree in aiming to do so without bringing in any nonphysical properties, instead appealing solely to the entirely physical, neural changes that ground the experience of seeing color.[7]

At this point, many antiphysicalists introduce a further line of argument. They insist that, if you "know what some experience is like,", then your introspective thinking about

7 For two such accounts, see Papineau (2002) and Montero (2007).

that experience will be manifestly "transparent," in that it will be guaranteed to reveal the essential properties of the experience. (When you think about an experience in terms of "what it is like," so the thought goes, isn't its real nature made apparent to you?) But such introspective thought does not show experiences to be physical states, point out the antiphysicalists, who conclude that they are not physical.

In response, physicalists will deny that the introspective mode of thinking about experiences is transparent in the relevant sense. While they maintain that experiences do have a physical nature, they say that there is no reason why introspective thinking should reveal this immediately. If physicalism is true, the mental is entirely grounded in inanimate aspects of the world. When we introspect, on the physicalist's view, we are aware of the mental but not aware of what grounds it.

Still, many feel that there is something very counterintuitive about a physicalist view of consciousness, and no doubt this lends plausibility to the antiphysicalist arguments. But an intuition is not an argument, and in any case physicalists have offered a number of suggestions to account for the intuitive difficulty of embracing physicalism (see Papineau 2011).

In the end, the strongest reason for distrusting an antiphysicalist view of consciousness is the strength of the contrary argument we have already examined in favor of physicalism. Somebody who denies physicalism about consciousness needs to find some fault in this argument, and the options are limited. We shall now examine these options.

The Causal Argument Analyzed

The defense of physicalism we have considered so far has come to be known in the philosophical literature as "the causal argument." To repeat, the general idea of the argument is this: a nonphysical mind would be like a ghost in a machine that has the power to flip switches and thereby cause our physical bodies to move. However, we have good reason to believe that all of these machine switches are flipped on or off by other physical parts of the machine. And since it is absurd to think that the switches are doubly flipped by both the machine and the ghost, we should conclude that there is no ghost in the machine, that the mental causes of our bodily movements are themselves physical parts of the machine. To help us explore the alternatives open for those who wish to resist physicalism, let us lay out the argument formally with three premises:

(1) Mental (and biological, etc.) states have physical effects.
(2) All physical effects have full physical causes.
(3) The physical effects of conscious causes aren't systematically overdetermined by two or more distinct causes.

Given these premises, the conclusion that the mental (biological, etc.) states mentioned in premise (1) must be part of the physical causes mentioned in premise (2) is inescapable. (After all, (1) tells us certain effects have mental (biological, etc.) causes, (2) tells us those effects have full physical causes, and (3) tells us those effects don't systematically have two or more distinct causes. The only way for these all to be true is for the mental (biological, etc.) causes to already be included in the physical causes.)

So someone who wants to deny physicalism about conscious mental states needs to deny one of the premises. What are the options?

Epiphenomenalism and Preestablished Harmony

A first option is to deny (1). You can hold that, despite first appearances, conscious mental states like pains, feelings, and decisions do not in fact cause bodily movements or any other effects in the physical world.

This position is most familiar in the guise of *epiphenomenalism*. Epiphenomenalists view the conscious mind as an inefficacious side-effect of the brain's operations. They agree with physicalists that bodily movements and so on are fully accounted for by brain processes, but insist that the conscious mind floats above these brain processes, as it were, rather like the puffs of smoke that are emitted by a steam train, but which themselves make no causal contribution to the train's progress.

An alternative way of denying (1) is to adopt Leibniz's doctrine of *preestablished harmony*. On this view, not only does the mind not affect the body, but the body doesn't affect the mind either. Rather, God has set both to run along parallel tracks, marching in perfect step so that we remain unaware of their causal independence.

The basic objection to both these doctrines is that they are theoretically quite implausible. Epihenomenalism requires us to suppose that conscious states, even though they are caused by processes in the physical world, have no effects on that world. This is a very odd kind of causal structure. Nature displays no other examples of such one-way causal intercourse between realms. Preestablished harmony is, if anything, even odder. Again, there are no other examples of nature dividing itself into causally isolated but coordinated realms. Normal principles of theory choice would seem to argue strongly against such convoluted causal structures, and in favor of the far simpler physicalist position that integrates the mental realm with the causal unfolding of the spatiotemporal world in an entirely familiar way.

If we are going to appeal to a principle of simplicity at this point, we might start wondering what work is being done by the causal argument laid out in the previous section, and in particular by its second causal closure premise. If general principles of theory choice can justify physicalism, why bring in all the complications associated with causal closure? The answer is that causal closure is needed to rule out interactionist dualism. General principles of theory choice may dismiss epiphenomenalism in favor of physicalism, but they do not similarly discredit interactionist dualism. As the brief historical sketch earlier will have made clear, interactionist dualism offers a perfectly straightforward theoretical option, requiring no commitment to any bizarre causal structures. Certainly, the historical norm has been to regard it as the default account of the causal role of the mental realm. Given this, arguments from theoretical simplicity are ineffective against interactionist dualism. Rather, the case against interactionist dualism hinges crucially on the empirical thesis that all physical effects have full physical causes. It is specifically this claim that makes it difficult to see how dualist states can make a causal difference to the physical world.

It is sometimes suggested that physicalism about the mind can be vindicated by an "inference to the best explanation." The thought here is that there are many well-established synchronic correlations between mental states and brain states, and that physicalism is a "better explanation" of these correlations than is epiphenomenalism (Hill 1991; Hill and McLaughlin 1999; Melnyk 2003). From the perspective outlined here, this starts the argument in the middle rather than the beginning, by simply assuming the relevant mind–brain correlations. This assumption of pervasive synchronic mind–brain correlations is only plausible if interactionist dualism has already been ruled out. After all, if we believed

interactionist dualism, then we would have no particular reason to think that every mental state is systematically correlated with some particular kind of brain state – since mental states, on the interactionist view, do not depend on any underlying neural processes.

Denying Causal Closure

These last points immediately indicate another possible response to the causal argument for physicalism. You could deny the second causal closure premise and embrace interactionist dualism. As observed earlier, there is nothing theoretically unattractive about this option. As far as initial plausibility goes, many see interactionism as preferable to physicalism. The case against it rests not on the superior plausibility of physicalism, but on the empirical case for causal closure.

It is noteworthy that almost no contemporary theorists defend interactive dualism. There were still a number of serious thinkers 50 years ago – the philosopher Karl Popper and the neuroscientist John Eccles spring to mind – who endorsed the view that nonphysical factors operated in the brain to influence its physical operations. But nowadays such views have few advocates in scientific or philosophical circles. We take this to testify to the strength of empirical evidence that built up through the 20th century against the existence of any nonphysical force fields. (Figures like Popper and Eccles were not so much eccentric as *old*: they grew into intellectual maturity at a time when mental and vital forces were widely taken for granted.)

It might occur to some readers that, if we are focusing on 20th-century science, then the indeterminism of modern quantum mechanics surely counts against the causal closure of the physical realm, and therewith undermines the causal argument for physicalism. Doesn't quantum mechanics show us that plenty of physical effects are chancy, and so don't have full physical causes? And doesn't this then leave room for an independent nonphysical mind to come in and affect what happens in the physical world?

This objection, however, is readily addressed. Even if quantum mechanics implies that some physical effects are themselves undetermined, it provides no reason to doubt a quantum version of the causal closure thesis, to the effect that the *chances* of those effects are fully fixed by prior physical circumstances. And this alone is enough to rule out any role for nonphysical causes. Such nonphysical causes, if they are to be genuinely efficacious, must make an independent difference to the chances of physical effects, and this in itself would be inconsistent with the quantum causal closure claim that such chances are already fixed by prior physical circumstances. Once more, it seems that anything that makes a difference to the physical realm must itself be physical.

It is striking that even those contemporary philosophers who are persuaded by the antiphysicalist arguments about consciousness do not typically respond by denying causal closure and embracing interactionism. Rather, they tend to go for epiphenomenalism.[8] Both Frank Jackson and David Chalmers initially combined their rejection of physicalism

8 But see Lowe (2000; 2003), who suggests that there may be "invisible" violations of causal closure, due to the irreducible mental causes of my bodily behavior always themselves having physical causes. On Lowe's view, while those bodily behaviors might *seem* to follow by physical law from the physical antecedents of their mental causes, they would not in fact ensue in possible worlds that shared all our physical laws but lacked the extra mental processes that occur in this world. Our response to such a position is, again, that science has failed to turn up any evidence for physical behaviors that cannot be fully accounted for in terms of prior physical processes alone.

with an advocacy of epihenomenalism (though it should be said that the difficulties of epiphenomenalism have since persuaded them to move away: Jackson is now a physicalist, while Chalmers favors a "neutral monism" whose differences from physicalism are a matter of debate).[9]

Denying No Overdetermination

One final way for nonphysicalists to evade the causal argument would be to deny premise (3), and so have the physical effects of conscious mental causes systematically caused twice over, both by a brain process and by an independently efficacious mental state. However, while this "belt and braces" option has had one or two defenders (Crane and Mellor 1990; Mellor 1995), it seems just as open to the accusation of unnecessary complexity as epiphenomenalism, if not more so. While occasional overdetermination by independent causes does sometimes occur (a man might be shot and struck by lightning simultaneously, for example), nature does not seem to offer any other examples of a whole *category* of effects that is *systematically* overdetermined by two independent causes. (Not to mention that the "belt and braces" view would seem to be in need of some additional mechanism to ensure that both causes are in place whenever one of them is.)

It is worth distinguishing this kind of "vicious" overdetermination from a species of "benign" overdetermination favored by some advocates of nonreductive physicalism. In order to explain this point, it will be necessary to backtrack a little.

Recall our initial characterization of physicalism. We were at pains to explain that it is a relatively permissive doctrine, and is not committed to the thesis that all properties, including mental properties, can be defined in the language of basic physics. Rather, the essential core was only the "metaphysical supervenience" of all the facts on the basic physical facts: once the basic physical facts are fixed, nothing more is needed for all the facts to be in place.

Let us use the term "*reductive* physicalism" for the stronger doctrine that all properties, including mental properties, can be identified with properties that are definable in the language of basic physics. Then "*non*reductive physicalism" is the weaker claim that all the facts supervene on the basic physical facts or are basic physical facts themselves. Such a view allows for "multiple realization"; that is, it allows for higher-level properties, including mental properties, to be composed by different lower-level physical properties in different cases. Pain, on this view, is not identical to a certain type of neural process, such as C-fiber stimulation, but can be determined by, or supervene on, different types of physical process in different kinds of creature. Such a view allows for extraterrestrials, if there are any, to experience pain while having very different underlying physical natures than we do.

Most contemporary physicalist philosophers probably favor nonreductive over reductive physicalism. But this, as some see it, generates a problem. If pain, for example, is not strictly identical to some lower-level physical cause such as C-fiber stimulation, then every time your C-fibers cause you to scream out and writhe, your pain would seem to act as another cause of your behavior, and thus it appears that the physical effects of mental causes are, after all, doubly caused. Thus, nonreductive physicalism might seem to saddle us with unacceptable proliferation of overdetermining causes for the physical effects of mental causes after all: both the physical cause implied by the causal closure thesis and the distinct mental cause.

9 See Montero (2015).

Advocates of nonreductive physicalism have a response: as they see it, there is nothing wrong with such an apparent duplication of causes if it is also specified that one cause metaphysically supervenes on the other. The issue here hinges on the acceptability of different kinds of overdetermination (Bennett 2003; Melnyk 2003). All can agree that it would be absurd if the physical effects of nonphysical causes always had two completely independent causes. However, even if such "vicious overdetermination" by two ontologically independent causes is so ruled out, as assumed by the causal argument, this does not necessarily preclude "benign overdetermination" by both a physical cause and a metaphysically supervenient mental cause, argue advocates of nonreductive physicalism. In their view, this kind of overdetermination is acceptable, on the grounds that the two causes are not ontologically distinct: the nonphysical cause isn't additional to the physical cause (nothing more is needed for your feelings than your brain states).

Not everybody agrees that this solution is satisfactory. A substantial minority among physicalists feel that the causal status of nonreduced mental categories is suspect, and so conclude that the only good physicalism is reductive physicalism. This is not the place to adjudicate this technical dispute.

The important point for present purposes is to note that this is an internal dispute among physicalists, rather than an argument against physicalism. It is agreed on all sides that the causal closure of the physical requires *at least* the metaphysical supervenience of the mental on the physical, on pain of denying any causal efficacy to mental states or positing a systematic vicious overdetermination. The only further question is whether it also requires the strong reduction of mental properties to physical ones, or whether it is acceptable to leave the mental properties as merely metaphysically supervenient.

As we said at the beginning, "naturalism" means different things to different people. But one central question raised by the term is whether mental, biological, and other such entities are all physical. We have tried to show how this strand in naturalism has strong scientific support.

References

Armstrong, D. (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Bennett, K. (2003). Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It. *Noûs* 37: 471–497.

Crane, T. and Mellor, D. (1990). There is No Question of Physicalism. *Mind* 99: 185–206.

Davidson, D. (1970). Mental Events. In *Experience and Theory*, edited by L. Foster and J. Swanson. London: Duckworth.

Elkana, Y. (1974). *The Discovery of the Conservation of Energy*. London: Hutchinson.

Feigl, H. (1958). The "Mental" and the "Physical." In *Minnesota Studies in the Philosophy of Science*, vol. II, edited by H. Feigl, M. Scriven, and G. Maxwell. Minneapolis: University of Minnesota Press.

Hill, C. (1991). *Sensations*. Cambridge: Cambridge University Press.

Hill, C. and McLaughlin, B. (1999). There are Fewer Things in Reality than are Dreamt of in Chalmers' Philosophy. *Philosophy and Phenomenological Research* 59: 445–454.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127–136.

Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.

Lewis, D. (1966). An Argument for the Identity Theory. *Journal of Philosophy* 63: 17–25.

Lowe, E.J. (2000). Causal Closure Principles and Emergentism. *Philosophy* 75: 571–585.

Lowe, E.J. (2003). Physical Causal Closure and the Invisibility of Mental Causation. In *Physicalism and Mental Causation*, edited by S. Walter and H.-D. Heckmann. Exeter: Imprint Academic.

Mellor, D. (1995). *The Facts of Causation*. London: Routledge.

Melnyk, A. (2003). *A Physicalist Manifesto*. Cambridge: Cambridge University Press.

Montero, B. (1999). The Body Problem. *Noûs* 33: 183–220.

Montero, B. (2006). What Does the Conservation of Energy Have to Do with Physicalism? *Dialectica* 60: 383–396.

Monetero, B. (2007). Physicalism Could Be True Even if Mary Learns Something New. *Philosophical Quarterly* 57: 176–189.

Montero, B. (2013). Must Physicalism Imply the Supervenience of the Mental on the Physical? *Journal of Philosophy* 110: 93–110.

Montero, B. (2015). Russellian Physicalism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*, edited by Torin Alter and Yujin Nagasawa. Oxford: Oxford University Press.

Nagel, T. (1974). What Is it Like to Be a Bat? *Philosophical Review* 82: 435–450.

Oppenheim, P. and Putnam, H. (1958). Unity of Science as a Working Hypothesis. In *Minnesota Studies in the Philosophy of Science*, vol. II, edited by H. Feigl, M. Scriven, and G. Maxwell. Minneapolis: University of Minnesota Press.

Papineau, D. (2001). The Rise of Physicalism. In *Physicalism and Its Discontents*, edited by B. Loewer and C. Gillett. Cambridge: Cambridge University Press.

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2011). What Exactly is the Explanatory Gap? *Philosophia* 39: 5–19.

Papineau, D. and Spurrett, D. (1999). A Note on the Completeness of "Physics." *Analysis* 59: 25–29.

Smart, J. (1959). Sensations and Brain Processes. *Philosophical Review* 68: 141–156.

Woolhouse, R. (1985). Leibniz's Reaction to Cartesian Interaction. *Proceedings of the Aristotelian Society* 86: 69–82.

14

Natural Mind

BRIAN L. KEELEY

Introduction

Naturalism with respect to a given domain, in general, is the belief that the tools and concepts of natural science are necessary to come to an understanding of the phenomena within that domain. Naturalism concerning the mind, then, is the belief that mental phenomena are amenable to natural scientific explanation, whereas antinaturalists concerning the mind believe that there is good reason to doubt that natural science is going to help us understand at least some crucially important aspects of mind.[1]

In his 1993 book, *Philosophical Naturalism*, David Papineau puts the issue this way:

> Philosophical problems are characterized by a special kind of difficulty, a difficulty which means that they cannot be solved, as scientific problems normally are, simply by the uncovering of further empirical evidence. Rather they require some conceptual unraveling, a careful unpicking of implicit ideas, often culminating in the rejection of assumptions we didn't realize we had. But still, despite these differences, there is clearly a sense in which philosophical thinking of this kind is part and parcel of the construction of scientific theories. Even if there is no direct involvement with empirical evidence, the task of the philosopher is to bring coherence and order to the total set of assumptions we use to explain the empirical world.

1 As I just noted, naturalism is a position that can be taken with respect to many domains, not only the mind. For example, one might advocate naturalism (or non-naturalism) concerning ethics, depending on whether one thinks science (and science alone) ought to be our guide to right and wrong. Or one might be a naturalist (or non-naturalist) concerning the origin of the world or the origin of life, depending on one's views about the nature of those things. One's stands on these views need not be connected; one might be a naturalist about the origin of life on earth but a non-naturalist about the mind. However, in this chapter, unless I note otherwise, I will restrict my discussion to the mind. In what follows, when I write "naturalist," you should understand it to be shorthand for "naturalist with respect to the mind." That said, I'll briefly discuss naturalism in a biological context in the next section.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

> The question at issue is whether *all* philosophical theorizing is of this kind. Naturalists will say that it is. Those with a more traditional attitude to philosophy will disagree. (1993, 3)

What Papineau observes for philosophy as a whole also applies to philosophical theorizing about the mind. An important starting point for this discussion is that we currently do not understand the mind; there are important aspects of mental phenomena the nature of which we currently do not grasp sufficiently. Once, we felt this way about many phenomena: What is lightning? How are rainbows produced? What is the moon made of? What is fire? and so on. Once, those questions were unanswered alongside these: Why do I find it harder to remember the names of people than their faces? How can I have a thought about something that doesn't exist, such as a unicorn or Sherlock Holmes' twin sister? What does it feel like to give birth? (When asked by a male.) Are the pains reported by human mothers during childbirth comparable to those felt, if they are felt, by females of species *x*?

Though there always seems to be more to learn about any phenomenon, we now feel that we grasp the nature of rainbows, lightning, the moon, and fire. However, there is much about the mind of which we are still ignorant. Indeed, as it once was with rainbows, lightning, and the moon, the very nature of mental phenomena – Are they nothing more than brain processes? If so, are they some specific kind of brain process? – seems to be currently up for grabs.

How one stands on the viability of a naturalist approach to understanding the mind is a kind of a bet on the future of our current ignorance about the mind. Naturalists are putting their money on the tools and concepts of science to dissolve our ignorance and bring us explanations and answers to these kinds of question about the mind. Non-naturalists advocate betting the other way: they argue that it is likely that science will not be able to answer the questions we have about the mind, or at least not all of them, or at least not the important ones.

Why bet on naturalism? One reason comes from a look at the past. The history of science exhibits the ongoing march of naturalism, with mystery after mystery being subsumed under the explanatory framework of natural science.

Let me demonstrate this claim by telling two stories. The first comes from Bertrand Russell's (1912) *Problems of Philosophy*. In the final chapter, "The Value of Philosophy," Russell is grappling with the conundrum that confronts any student of philosophy: The sciences can point to increased knowledge and ability to control the world as results, and on that basis science has a pretty clear source of value. Philosophy, on the other hand, is famous for having *not* answered its central questions. That being the case, where can philosophy look for a source of value? To answer this question, Russell provides a very quick characterization of philosophy and its relationship to other areas of study:

> Philosophy, like all other studies, aims primarily at knowledge. The knowledge it aims at is the kind of knowledge which gives unity and system to the body of the sciences, and the kind which results from a critical examination of the grounds of our convictions, prejudices, and beliefs. But it cannot be maintained that philosophy has had any very great measure of success in its attempts to provide definite answers to its questions. If you ask a mathematician, a mineralogist, a historian, or any other man of learning, what definite body of truths has been ascertained by his science, his answer will last as long as you are willing to listen. But if you put the same question to a philosopher, he will, if he is candid, have to confess that his study has not achieved positive results such as have been achieved by other sciences. It is true that this is partly accounted for by the fact that, as soon as definite knowledge concerning any subject becomes possible, this subject ceases to be called philosophy, and becomes a separate science.

> The whole study of the heavens, which now belongs to astronomy, was once included in philosophy; Newton's great work was called "the mathematical principles of natural philosophy."…Thus, to a great extent, the uncertainty of philosophy is more apparent than real: those questions which are already capable of definite answers are placed in the sciences, while those only to which, at present, no definite answer can be given, remain to form the residue which is called philosophy. (1912, 154–155)

Russell's discussion here involves a reading of the history of Western philosophy: In the beginning of inquiry, all was philosophy, because we did not have systematic knowledge, and more importantly, we didn't have a systematic way of coming to knowledge. We did not have a reliable epistemology. But slowly, we, as a species, began to learn about the world as we *learned how to learn* about the world. Areas of inquiry split off from philosophy and became established as disciplines of their own. Astronomy was one of the earliest fields to be established, but it was soon joined by other parts of what we now know as physics: mechanics, optics, and so on. As the centuries reeled out, physiology, anatomy, botany, the biological sciences, chemistry, geology, and the rest of modern science ceased to be the stuff of philosophical speculation and became distinct fields of study. As a result, at any given time in its history, philosophy is the repository of those questions to which we currently do not have answers; indeed, we do not even have a clear idea *how* to go about answering them. As soon as we do, a new science is born.[2]

My second story derives from a sentence I elided from the same long quotation from Russell. Where the ellipsis appears, Russell writes, "Similarly, the study of the human mind, which was a part of philosophy, has now been separated from philosophy and has become the science of psychology." In 1912, when Russell published his book, this was a relatively new development in the intellectual history of the Western world. However, this sentence passes too quickly over the history relevant to our concern with naturalism about the mind, so let me spell out a more full story leading up to Russell's observation here, because it will allow us to more clearly articulate what is at issue for us.

If naturalism is defined in terms of the techniques and concepts of natural science, there's a sense in which it cannot exist prior to the advent of natural science itself. For this reason, a more careful retelling of my Russellian story of the history of philosophy would need to speak of the development of *protosciences* in the centuries before the beginning of the modern era (in roughly the 16th century, when Copernicus developed his heliocentric model, and the early 17th century, with the work of Galileo and Harvey). It was during this period when the polymath René Descartes made his contributions in mathematics, physiology, and, most famously, philosophy. Descartes is a key figure in the history of studies of the mind, as well as of the natural sciences more generally. Descartes was clearly very interested and involved in the embryonic sciences of his day. Further, he was no doubt well aware of the growing conflict between these developing ideas of science and the authority of the Christian Church. The Church was a powerful institution of the day, and developments in science – especially concerning the structure of the solar system, but also in other areas – were drawing negative attention to scientists and their ideas.

2 As stories go, Russell's is interesting, but it is not without its complications. For example, geometry and the rest of mathematics clearly made an early break from philosophy. Some would argue that logic has slowly been establishing its separate existence for a little more than a century now. Neither of these formal endeavors are "sciences," in the normal sense of that term. Similarly, theology has separated from philosophy as well. Obviously, to tell a full story of these developments, Russell would need more than a single paragraph!

According to Descartes, the world is divided into two independent substances. On the one hand, there is *material* substance, divisible (consisting of parts) and extended (taking up, and located in, space). On the other hand, there is immaterial substance (including *mental* substance), indivisible and non-extended. Dividing the world this way also creates separate and non-overlapping magisteria[3] for the Church and science. The study of matter – its composition and properties, its movement and transformations – is the domain of science. The mind – a.k.a. the "soul," the locus of free will, and that which might well be immortal and outlast the death and dissolution of the (material) body – is the domain of religious authority. Against the backdrop of the tensions between science and religion, Descartes' metaphysical dualism can be understood as a grand compromise between a powerful Church and a nascent natural science.

However, while it can be argued that Descartes' compromise cleared the way for the next few centuries of developments in the physical and even biological sciences, there was a concomitant, negative effect on the development of the science of the mind. The Cartesian mind, standing outside of the natural order of material substance, was not an appropriate object of study by science, as Descartes and many of his intellectual descendants thought of it. It is perhaps for this reason that we only see psychology as a natural science splitting away from philosophy of mind in the mid-19th century. Wilhelm Wundt is generally given credit for founding the first experimental psychology laboratory at the University of Leipzig in 1879. Germany was the locus of much early scientific work on the nature of mind, including that of Wundt's mentor, Hermann von Helmholtz, who made important contributions in sensory psychology, and Herman Ebbinghaus, who conducted early work on learning and memory. During this same period, American labs were being set up, both by Wundt's student G. Stanley Hall, the first president of the American Psychological Association, who studied child development and education at Johns Hopkins University, and by the philosopher/psychologist William James at Harvard, who in addition to making important contributions to the theory of emotions and the self, also published *Principles of Psychology* (1890/1981), a landmark English-language work of scientific psychology. In Russia, Ivan Pavlov was conducting his experiments on conditioning as a form of learning and memory, receiving a Nobel Prize for his efforts in 1904. It is against this background, and that of other work that there is not space to discuss here, that Russell writes, "the study of the human mind…has become the science of psychology" (1912, 155).

Notice, however, that just as Russell tells a story of the history of philosophy writ large as the splitting off of portions of it over time to form other fields, leading to its slowly shrinking, one can focus in and tell a very similar story about the history just of philosophy of mind. In the beginning, all questions of mind were part of philosophy. Indeed, one might say that, thanks to Descartes, they were *necessarily* outside of the purview of natural science. But starting roughly in the 19th century, with the birth of experimental psychology, we started learning how to systematically study the mind using the tools and concepts of natural science. Aspects of the mind began to be naturalized. As mentioned in the previous paragraph, the aspects of mind that first became amenable to study included the senses, emotion, child development, learning, and memory. That did not spell the end of philosophy of mind, because it left many aspects unexplained, such as the nature of consciousness, the problem of other minds, free will, and the nature of rationality. Also, what is known as the mind–body problem – the ultimate nature of what we call the "mind"

3 To borrow a term from Stephen Jay Gould (1997).

and the "body" (or "brain") and their relationship – was largely left untouched by this early experimental psychology. So, part of the "residue" left as philosophy is questions in the philosophy of mind.

Returning then to the central issue of this chapter, I can rephrase things this way: Given that some, arguably important, aspects of mind have already been naturalized (i.e., become the purview of experimental psychology – an established and robust branch of natural science), what of the rest of the mind? The naturalist looks at the last 200 years of advances in the sciences of the mind and, based on the successes so far – I didn't even describe the wealth of developments in the 20th century, including the birth of cognitive psychology, cognitive science, and consciousness studies – is willing to bet that such development is bound to continue. The non-naturalist is nowhere near as sanguine, betting instead that experimental psychology and the rest of the cognitive sciences have peeled off the easy problems of mind amenable to naturalistic explanation, but have left behind a hard core of unexplainable (by natural science, at least) issues. If the philosophy of mind is a vegetable or fruit, the naturalist takes it to be an *onion*, already half-peeled and eventually to be peeled away entirely, with no remainder. The non-naturalist, to the contrary, sees philosophy of mind as a *mango*; the outer skin and flesh is there to be carved away, but eventually you work your way down to a hard, inedible stone at the center.

Methodology vs. Metaphysics

One place where discussions of naturalism have been brisk in recent years is in debates over the origin of species in biology. There are those, such as Phillip E. Johnson (2001), who argue in favor of an "intelligent design" account based on an understanding of why we see the organisms on earth that we do. He argues that our best explanation of the diversity of life found on our planet is that it has been guided by the actions of some intentional and intelligent agency. Johnson and his fellow travelers are generally careful not to equate this intelligent designer with the Judeo-Christian-Islamic God, as they typically focus on the need to establish some kind of *design* rather than saying much about the *designer*. However, they generally are clear that the role of this designer in the development of biology is non-natural. Indeed, Johnson is explicit that the rejection of naturalism in biology is an essential first step in understanding the biological world and that it is the *a priori* commitment to what he terms "philosophical naturalism" that prevents "Darwinists" from being open to the intelligent design hypothesis. Johnson's concept of philosophical naturalism is a view according to which supernatural forces cannot be invoked as part of any explanation of the world described by science, and Johnson believes that this prior philosophical commitment of his opponents inappropriately stacks the deck against intelligent design. Further, he claims that, "If creation is admitted as a serious possibility, Darwinism cannot win, and if it is excluded *a priori* Darwinism cannot lose" (2001, 66). For Johnson, philosophical naturalism is the linchpin in debates over the origin of species.

Discussing this application of naturalism to debates over the origin of biological diversity would take us too far from the topic of this chapter, but I raise the example because it allows me to introduce a useful distinction made by one of Johnson's detractors. In a rebuttal to Johnson's arguments against philosophical naturalism, Robert Pennock (2001) argues that Johnson's critique fails, in part, due to a conflation within his discussion of naturalism.

After complaining that Johnson could have spent more time spelling out explicitly what he means by his use of the term "naturalism," Pennock goes on to argue that he seems to be confusing two different senses of it, one of which would be damning if opponents of intelligent design held it (but they don't), and one of which those opponents do indeed hold (but which is far more innocuous).

As Pennock presents it, one sense of naturalism "makes a commitment to substantive claims about what exists in nature, and then adds a closure clause stating 'and that is all that there is'" (2001, 84). He calls this view *ontological naturalism* and notes that such a metaphysical commitment has been made more than once in Western history, such as when 17th-century mechanistic materialists proposed that the world consisted of nothing but material particles in motion (this ran contrary to the Cartesian dualism discussed in the previous section). This form of naturalism would have the effect that Johnson is worried about: if God is taken to be supernatural and one had a prior commitment to a form of ontological naturalism that precluded any supernatural agency, then one would have a situation in which a role for a supernatural designer was ruled out *a priori*.

However, Pennock holds that contemporary proponents of evolutionary biology have no such philosophical commitment. Instead, they maintain what he terms *methodological naturalism*:

> The Methodological Naturalist does not make a commitment directly to a picture of what exists in the world, but rather to a set of methods as a reliable way to find out about the world – typically the methods of the natural sciences, and perhaps extensions that are continuous with them – and indirectly to what those methods discover. An important feature of science is that its conclusions are defeasible on the basis of new evidence, so whatever tentative substantive claims a Methodological Naturalist makes are always open to revision or abandonment on the basis of new, countervailing evidence. Because the base commitment of a Methodological Naturalist is to a mode of investigation that is good for finding out about the empirical world, even the specific methods themselves are open to change and improvement; science may adopt promising new methods and refine existing ones if doing so would provide better evidential warrant. (2001, 84–85)

It is to this latter form of naturalism that Pennock believes contemporary evolutionary biologists and naturalist philosophers of biology are committed. Further, he argues that it does not rule out an intelligent design hypothesis *a priori*, as Johnson complains. Instead, it puts the ball in the creationists' court and challenges them to show how explaining the diversity of species on earth requires invoking some overarching purpose or intention on behalf of a guiding agency, the proposed intelligent designer. Given the in-principle defeasibility of the current account of evolution deriving from Darwin's work, it is open to intelligent design's opponents to demonstrate how their explanatory approach is superior. Pennock goes on to complain that such a positive story isn't what he finds in Johnson's work; instead, he claims that it is overwhelmingly negative in structure, attempting to show that "Darwinists" are deceptive or incorrect in their proposals.

Again, my purpose here is not to review the debates over intelligent design. Instead, I'm interested in making use of Pennock's distinction in the context of naturalism with respect to the mind. As with biology, I propose that it is useful to distinguish ontological from methodological naturalism in the philosophy of mind. However, there's no guarantee that this application will result in the same sorts of conclusion as Pennock proposes in his domain of naturalism.

An ontological naturalist about the mind would be one who argues that there are only certain things that the mind can be. For example, it is possible to read *logical behaviorism*, as proposed by Gilbert Ryle, as a kind of ontological naturalism, in that he proposes that all mental phenomena are to be understood in terms either of overt behavior or of dispositions to overt behavior. In *The Concept of Mind*, Ryle (1949) provides a behaviorist analysis of knowledge, the will, emotion, perception, imagination, and the intellect (to take only major chapter headings). Ryle finds himself in this position partly due to his reaction against the Cartesian dualism described earlier. As Ryle sets it up, if we wish to explain mental phenomena, we are faced with a dilemma. The first horn involves explanation by appeal to occult causes, such as essentially private motives, feelings, cogitations, desires, and so on. This is the "two worlds" explanatory strategy of metaphysical dualism. Dualism explains the public behavior we see when a person conducts a bodily act by referring to a second, occult world of the mind. Ryle's second horn is logical behaviorism: we can explain mental phenomena by appealing not to occult, immaterial causes, but to the overt, *public behavior* of agents; that is, the kind of behavior with which we are all familiar.

Ryle is motivated in this approach in large part by his rejection of Cartesian dualism, a view that attempts to explain mental phenomena by invoking a nonphysical mind: a "ghost in the machine," as Ryle (in)famously puts it. In rejecting the Cartesian mind as "spooky" and ghostlike, a logical behaviorist seems to be invoking ontological naturalism. There is something illegitimate about the mind of Cartesian or metaphysical dualism, and on those grounds the entire approach is wrongheaded and ought to be avoided.[4]

The philosophical position that immediately succeeded logical behaviorism – *mind–brain identity theory* – might also be read as continuing this ontological naturalistic move. Identity theorists were also suspicious of Cartesian dualism. However, instead of saying that the mind is behavior and dispositions to behavior (and nothing more), they argued that minds are brains (and nothing more). Where logical behaviorism seeks to reduce the mental to behavior, identity theory seeks to reduce the mental to a combination of behavior and the physical states of an organism. The early identity theorist Herbert Feigl writes:

> The solution [to the mind–body problem] that appears most plausible to me, and that is entirely consistent with a thoroughgoing naturalism, is an *identity theory* of the mental and the physical, as follows: Certain neurophysiological terms denote (refer to) the very same events that are also denoted (referred to) by certain phenomenal terms. The identification of the objects of this twofold reference is of course logically contingent, although it constitutes a very fundamental feature of our world as we have come to conceive it in the modern scientific outlook. Utilising Frege's distinction between *Sinn* ("meaning," "sense," "intension") and *Bedeutung* ("referent," "*denotatum*," "extension"), we may say that neurophysiological terms and the corresponding phenomenal terms, though widely differing in sense, and hence in the modes of confirmation of statements containing them, do have identical referents. I take these referents to be the immediately experienced qualities ["raw feels"], or their configuration in the various phenomenal fields. (1960, 38; emphasis in original)

4 Note also that Ryle's naturalism also resonates with Papineau's characterization of non-naturalists as "those with a more *traditional* attitude to philosophy" (1993, 3; my emphasis). Ryle's work, together with earlier work by the American pragmatists (see Fisch 1951, 19–20), ushered in a period of antagonism towards Descartes' philosophy of mind. Ryle sees Cartesian dualism as the received position within philosophy of mind and a traditional way of thinking that needed to be unseated. Indeed, I venture to guess that when Ryle is taught these days, it is more his rejection of Descartes than his positive, behaviorist account of the mind that receives the class's attention.

By logically identifying mental phenomena, such as being in pain, with neural phenomena, such as the activity of C-fibers in the peripheral nervous system, identity theorists can be seen as proposing that there is no independent ontological category of the mental. It was common in identity theorist discussions to compare the identity of mind and brain to other landmark discoveries in the history of science, such as that lightning is (and is nothing more than) atmospheric electric discharge or that temperature is (and is nothing more than) the vibration of molecules (see, e.g., Place 1956).

So, it is possible to hold ontological naturalism concerning the mind. To do so would be to restrict the kinds of entity one is willing to posit to explain mental phenomena. Pennock's other sense of naturalism – methodological naturalism – also represents a viable, but different, philosophical framework for understanding the mind. With methodological naturalism, one makes a commitment to a set of methods for studying mental phenomena. These are the methods of natural science, empirical investigation, testing of defeasible hypotheses, and so on.

Just as one can read logical behaviorism as hewing to the ontological naturalist line, so one can understand what is sometimes called *psychological behaviorism* as methodologically naturalist. It is a historical curiosity that at roughly the same time as philosophers such as Ryle were championing behaviorism in philosophy of mind, there were scientific psychologists, chiefly J.B. Watson and B.F. Skinner, who were advocating behaviorism as the correct approach to psychology. First Watson and then, later, Skinner built on Pavlov's previous work on learning and memory, developing it into a robust theory of human and non-human animal behavior. In doing so, they advocated that, as a science, psychology needed to stick to what it could reliably measure: stimuli, behavioral responses, and the manipulable relationship between the two. The two behaviorisms – logical and psychological – were subtly different, even beyond the different domains they addressed (philosophy of mind and scientific psychology, respectively). One difference is adverted to by Ron Amundson when he observes that, "Unlike psychological behaviorists, who embraced causal theories of behavior while rejecting mentalist vocabulary, logical behaviorists embraced the vocabulary, while denying the relevance of causal theories intended to explain the patterns of behavior addressed by mentalism" (1988, 505). In other words, psychological behaviorism was driven by methodological concerns, by what it could establish by empirical study; and that was the causal relationship between stimulus and response. Logical behaviorism, on the other hand, was more interested in metaphysics; that is, in what minds were, ultimately: behavior. Therefore, in Pennock's terms, both forms of behaviorism were naturalist, but one was more driven by ontological naturalism and the other by methodological naturalism.

Eliminative Materialism as a Naturalized Approach to the Mind

Next, just as there were both methodologically and ontologically naturalist approaches to behaviorism, there is a more methodological counterpart to mind–brain identity theory, a philosophical position known as *eliminative materialism*. Because, unlike most of the positions discussed so far, it is still a "live" option, in that there are a number of contemporary philosophers (including myself) who are sympathetic to it, I will spend a little more space explaining it.

Eliminative materialism shares with mind–brain identity theory the central idea that neuroscience is a key piece in the puzzle of unraveling the nature of the mind. However,

where identity theory derives from a metaphysical claim (mind = brain), eliminative materialism is motivated by explanatory concerns, chiefly that as science learns more about the brain, our understanding of ourselves and our minds will be greatly transformed, even in ways that might defy common sense. However, to see what eliminative materialists have in mind and how they get there, it helps to start with the work of Wilfrid Sellars, a philosopher who did not identify with the label "eliminative materialist," though one of his students, Paul M. Churchland, has practically become the poster boy for the view.[5]

Sellars distinguishes two "images," or very general philosophical frameworks for understanding human activity. The first is the *manifest image*: the embodiment of our commonsense understanding of human behavior, including our own personal behavior. Sellars characterizes "the manifest image of man-in-the-world as the framework in terms of which man encountered himself – which is, of course, when he came to be man" (1960/1963, 6). This image is not pretheoretical in the sense of being unreflective. Rather, it is the image of oneself achieved upon taking oneself as an object of understanding: what humans got when they first realized that they, too, were something that required understanding, in addition to all of the other confusing aspects of the world, including other animals, the weather, the night sky, and so on. Furthermore, it is a framework in which the basic ontological category is that of "persons." The manifest image emphasizes that everything understood is understood in terms of being a kind of person.

The manifest image is contrasted with what Sellars calls the *scientific image*. This is the image of our self and the world provided by the explicit theorizing of post-Enlightenment science. There is a strong "what-you-see-is-what-you-get" quality to Sellars' conception of the manifest image. He cites Mill's inductive method as central to the method of the manifest image; such explanation is generated by noting the correlations of observed events in the world (1960/1963, 7). In contrast, what demarcates the method of the scientific image is its method of hypothesis and the postulation of that which is unobserved or unobservable in the service of explanation. The fundamental ontology of the manifest image (persons) is directly observable to everyone; indeed, if all one had was the manifest image, persons are all one would ever see when looking at other humans. By contrast, the fundamental ontology of the scientific image – say, that provided by contemporary physics – is one of unobservable atomic elements, atomic forces, neural circuits, brain chemistry, and the like.

What is the relationship between these two images? They are often taken to be opposed to one another. As one striking example, one line of thought derives from taking the scientific perspective on humans themselves and seeing them not as persons in the sense of the manifest image but rather as a collection of abstract, scientific entities (cell assemblies, molecules, expressed DNA, quarks, what have you): "Even persons, it is said (mistakenly, I believe), are being 'depersonalized' by the advance of the scientific point of view" (1960/1963, 10). This is "mistaken" because Sellars takes the goal of philosophy to be explanation in the broadest sense: he sees both images as essential to a full understanding of humans, the world, and the place of humans in the world. He likens the relationship between the two to that of the different component images of a stereoscopic diagram. Properly viewed through a pair of stereoscopic lenses, the two images combine to provide a single image with dimensions lacking in either component on its own. Sellars' notion of these two different images of ourselves and the world around us shows up in what have come to be known as "folk theories."

5 For more on Paul Churchland and eliminative materialism, see Keeley (2005).

Folk theories are the common-sense theories possessed by the average person: the "folk." Specifically, eliminative materialists are concerned with a particular subset of folk theories: folk psychology. "Folk psychology" has different senses,[6] but here I am using it to refer to "the [everyday] conceptual framework of explanations of human behavior," explanations that are typically in terms of "beliefs, desires, intentions, expectations, preferences, hopes, fears, and so on" (Baker 1999, 319). I am deploying my folk psychology when I predict (or, afterwards, explain) that my friend will be standing in front of the movie theater at 7:50 next Tuesday by positing that he *believes* I will meet him there and he *desires* to keep our date and *expects* me to be there as well.

The important thing that eliminative materialism stresses about traditional folk psychology is that, as a theory, it is potentially false. Theories propose an explanation, and this proposal is risky. It is therefore, at least in principle, open to falsification, just as important elements of other folk theories, such as those found in folk physics (heavier objects fall faster than lighter ones) and folk biology (the spontaneous generation of mice from grain) have been shown to be false. Therefore, eliminative materialists argue that our folk psychological explanations of ourselves in terms of beliefs and desires may well turn out to be false.[7]

This potential falsification of our folk understanding of ourselves is a result of the *continuity* that naturalists see between the scientific image and manifest image. As Papineau proposes in the quotation with which I opened this chapter, naturalists see philosophy and the sciences as doing different things, but these projects are nonetheless continuous with one another. Or, if continuity is not the correct metaphor here, the two images need to be structured such that they can be profitably *interpolated*, as suggested by Sellars' metaphor of combining the two images as in a stereoscope, where neither on its own is capable of giving us the full story.

Take, for one example, the need for and nature of sleep.[8] Common sense would suggest that sleep is needed for rest, to regain our vigor after a hard day's work. But science has shown that mere rest alone is not sufficient, as relaxation without falling into sleep leads to an unhealthy state. Instead, neuroscience is indicating more and more that sleep is a complex brain state required for healthy neurochemistry. Over the course of a day, the normal activity of mammalian brains builds up biochemical imbalances that can be reset only by changing the brain state. That recovery process is what we experience as tiredness leading to sleep. A full, naturalist understanding of the phenomenon of sleep would involve both a scientific account of its role in physiology and a more manifest image of what it feels like to sleep, dream, and wake. Folk psychology might give us a starting point – a source of hypotheses – concerning sleep, but it should not be the end of the story.

Or, to take another example, consider the phenomenon of feeling that we are the "owners" of our agency.[9] When I move my arm, the act of moving feels as though it is me that is bringing about the movement; having the armed moved up by a puppet string feels importantly different. Feeling that I am the agent that moves my own body is clearly central

6 See Ravenscroft (2010).

7 See the work of Paul Churchland, starting with (1981/1989), for a forceful discussion of the empirical foibles of folk psychology. However, it is not a given that our folk understanding of ourselves should be seen as a theory (and hence as something falsifiable). This is a substantive and controversial claim of eliminative materialism, though it is a point that does not get a lot of attention these days. For an opposing view, see Wilkes (1984, 1991).

8 An example raised briefly in Churchland (1981/1989).

9 This example is taken from Bayne (2011).

to my common-sense image of myself, but how to understand that feeling and its relationship to other aspects of my mind is not obvious. Should it be thought of as a kind of perception? Do I perceive myself to be acting? Or is it less perceptual than conceptual; that is, do I instead think myself to be acting? Or, as a third alternative, is it a mental state fully on the motor side of things, such that just as the use of the eyes or the ears has characteristic experiences associated with it (called "perceptions"), so too does the activity of the motor cortex in the generation of a bodily movement have its own characteristic experience? Are these folk psychological categories of *perception*, *thinking*, and *motor action* even the correct categories to be used when confronted with a neuropsychiatric phenomenon such as *alien hand syndrome*, in which an afflicted person finds their hand moving, as if of its own volition, without feeling that it is he or she that is responsible for that movement? Here's a case were folk psychology does not give unambiguous answers and we may need to look to science to help us fully understand what is going on when we feel ourselves act.

Will eliminative materialism turn out to be the best strategy for understanding the mind? Who knows? As I mentioned at the beginning of this chapter, the naturalist is one who is willing to make a bet that the tools and concepts of science will turn out to be essential to the process of understanding the mind. Or, if you do not like the gambling metaphor, think of naturalist approaches to the mind, as exemplified by eliminative materialism, as a kind of working hypothesis.

Conclusion

I should begin this conclusion by acknowledging that my exploration of naturalism concerning the mind has been somewhat unorthodox, and not just in the anecdotal and historical approach I take. There are other, perhaps more direct ways in which one can go about it. For example, one might make the Darwinian point – most explicitly pursued in his *Descent of Man* (1871) and *Expression of Emotions in Man and Animals* (1872), but also skillfully taken up by his disciple George Romanes (1883) – that we humans are evolved animals alongside the rest of the animal kingdom. One implication of this continuity of humans with other animals is a continuity in our explanations of both. Given that we tend to be thoroughgoing naturalists with respect to all aspects of non-human animals, a similar naturalism is implied with us. We are not separate creations; as such, we do not call for separate explanations.[10]

Another related implication of the Darwinian point is that the evolutionary connection between all organisms means that we also share an evolutionary history of our minds. That is, Darwinism holds that minds themselves have evolved. To the extent that evolution is a naturalist-friendly phenomenon – setting aside the claims of Johnson and the intelligent design movement – mental phenomena must be amenable to naturalist exploration. One worry about a non-naturalist conception of the mind is that it is difficult (to put it mildly) to see how it might have evolved. However, by all appearances, it *did* evolve: selection forces acted on our hominid ancestors to bring about the kinds of mind we now have.

Another approach I could have taken might have been to explore the contemporary sciences of the mind/brain and show how in recent years it seems that all of the major questions of philosophy of mind, from consciousness to free will to moral reasoning to how

10 I explore some of these issues in Keeley (2004).

infants learn to meaningfully engage with the world have been assaulted on all sides by empirical exploration. If contemporary cognitive scientists are successful, there will be no "residue" of philosophy of mind remaining to be explained. The recent work of Patricia Churchland (2002; 2011; 2013) has taken this approach, though see Noë (2009) for a countervailing view.

Instead, in this chapter I aimed to explore and to differentiate different kinds of naturalist approach to studying the mind, as exemplified by a number of 20th-century movements within philosophy of mind. The latest of these, eliminative materialism, stresses the continuity of our scientific image of ourselves with our more common-sense understanding. That is, in other words, a proposal that our human minds have a naturalist side to them amenable to empirical exploration and explanation, as with any other part of the world. That is not to say that there is not a role for philosophical investigation as well, just so long as we realize that to understand our natural minds, *both* approaches are part of a general attempt to come to grips with the phenomena of the world.

Acknowledgments

I give my thanks to Ahmed Alwishah and Kelly James Clark for their comments and input on an earlier draft of this chapter.

References

Amundson, R. (1988). Logical Adaptationism. *Behavioral and Brain Sciences* 11: 505–506.

Baker, L. (1999). Folk Psychology. In *MIT Encyclopedia of Cognitive Science*, edited by Robert A. Wilson and Frank C. Keil, pp. 319–320. Cambridge, MA: MIT Press.

Bayne, T. (2011). The Sense of Agency. In *The Senses: Classic and Contemporary Philosophical Perspectives*, edited by Fiona Macpherson, pp. 355–374. Oxford: Oxford University Press.

Churchland, P.M. (1981/1989). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy* 78(2): 67–90. Reprinted as Chapter 1 of Churchland, P.M. (1989). *A Neurocomputational Perspective: The Nature of Mind and Structure of Science*. Cambridge, MA: MIT Press.

Churchland, P.S. (2002). *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P.S. (2011). *Braintrust: What Neuroscience Tells Us About Morality*. Princeton: Princeton University Press.

Churchland, P.S. (2013). *Touching a Nerve: The Self as Brain*. New York: W.W. Norton.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.

Feigl, H. (1960). Mind-Body, Not a Pseudo-Problem. In *Dimensions of Mind: A Symposium*, edited by S. Hook, pp. 24–36. New York: New York University Press.

Fisch, M.H. (1951). *Classic American Philosophers*. Englewood Cliffs: Prentice Hall.

Gould, S.J. (1997). Nonoverlapping Magisteria. *Natural History* 106: 16–22. Reprinted in Gould, S.J. (1998). *Leonardo's Mountain of Clams and the Diet of Worms*. New York: Harmony Books.

James, W. (1890/1981). *The Principles of Psychology*. Cambridge, MA: Harvard University Press.

Johnson, P.E. (2001). Evolution as Dogma: The Establishment of Naturalism. In *Intelligent Design and its Critics: Philosophical, Theological and Scientific Perspectives*, edited by Robert T. Pennock. Cambridge, MA: MIT Press.

Keeley, B.L. (2004). Anthropomorphism, Primatomorphism, Mammalomorphism: Understanding Cross-Species Comparisons. *Biology & Philosophy* 19: 521–540.

Keeley, B.L. (ed.) (2005). *Paul Churchland*. Contemporary Philosophy in Focus. New York: Cambridge University Press.

Noë, A. (2009). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. New York: Hill and Wang.

Papineau, D. (1993). *Philosophical Naturalism*. Oxford: Blackwell.

Pennock, R.T. (2001). Naturalism, Evidence, and Creationism: The Case of Phillip Johnson. In *Intelligent Design and its Critics: Philosophical, Theological and Scientific Perspectives*, pp. 77–97. Cambridge, MA: MIT Press.

Place, U.T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology* 47: 44–50. Reprinted in Place, U.T., Valentine, E.R., and Graham, G. (2003). *Identifying the Mind: Selected Papers of U.T. Place*. Oxford: Oxford University Press.

Ravenscroft, I. (2004). Folk Psychology as a Theory. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/fall2010/entries/folkpsych-theory/ (last accessed July 15, 2015).

Romanes, G. (1883). *Animal Intelligence*. Boston: Appleton.

Russell, B. (1912). *The Problems of Philosophy*. London: Oxford University Press.

Ryle, G. (1949). *The Concept of Mind*. Chicago: University of Chicago Press.

Sellars, W. (1960/1963). Philosophy and the Scientific Image of Man. In *Science, Perception and Reality*, edited by Wilfrid Sellars, pp. 1–40. Atascadero: Ridgeview.

Wilkes, K.V. (1984). Pragmatics in Science and Theory in Common Sense. *Inquiry* 27: 339–361.

Wilkes, K.V. (1991). The Relationship between Scientific Psychology and Common-Sense Psychology. *Synthese* 89: 15–39.

15

Naturalism and Dualism

JOSEPH LEVINE

Introduction

Dualism is a doctrine in the philosophy of mind, usually opposed to materialism (or physicalism; I will use these two terms interchangeably), though also to idealism. Materialism is the view that, in the relevant sense, nothing – in particular, neither the mind nor mental phenomena – is "over and above" the physical. What is meant by "over and above" is controversial, as is what is meant by "physical." I will address these issues shortly. Idealism is the view that, in the relevant sense, everything is ultimately mental in nature. This view is standardly associated with the empiricist philosopher Berkeley. Both materialism and idealism are versions of "monism," the view that everything is ultimately of one kind of stuff: physical, according to the materialist, and mental, according to the idealist. Dualism, on the other hand, is the doctrine that mind and matter are of very different kinds of stuff; reality breaks down ultimately into two, hence "dualism."

Naturalism is a doctrine in metaphysics and philosophy of science. While there are differences over the content of dualism and materialism, these are relatively well-defined terms compared to "naturalism." Philosophers intend a variety of doctrines by this term, as the essays in this volume testify. The precise content depends to a large extent on what it is being contrasted with. Often the contrast is with theism, the doctrine of most religions: that there is a God (or gods) who created the world and whose decisions and behavior are responsible for (a lot of) what happens in it. But sometimes the contrast is with dualism or related doctrines. I do not claim there is one right way to characterize naturalism. I think there are legitimate divergent uses of the term. However, I do think there is a core idea that underlies a lot of the different uses, and it's instructive to see how two different interpretations of this core idea give rise to the two specific naturalistic doctrines that contrast with dualism, on the one hand, and with theism, on the other. My task in this chapter is to draw out both the common underlying idea and the distinction it gives rise to.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

I will briefly characterize the core idea, before explaining how this is expressed in the sense of "naturalism" that contrasts with dualism – I'll call it "narrow naturalism" – and describing the debate between dualists and those who believe that the mind can be "naturalized" (the naturalists; that is, in this case, materialists).[1] Following that, I will show how interpreting the core idea differently gives rise to the contrast with theism. On this interpretation, dualists can be naturalists as well. Let's call this sense of naturalism "broad naturalism." Finally, I will conclude with a brief defense of the claim that dualism is quite compatible with broad naturalism against an argument of Robert Adams (1987).

The Core Idea

When most people are asked what they mean by the "natural," they typically respond by contrasting it with phenomena that involve, or are the product of, mental or intentional activity. For many people, the main term that contrasts with "natural" is "artificial." So, what makes a food ingredient natural is that it's not artificial, where the latter is explicitly humanly produced, while the former is found in nature without human intervention. The core idea, then, is the contrast between the mental and the non-mental, the latter being the natural.

If the natural is defined as the non-mental, then we need to have a more positive characterization of the mental: it can't just be the "unnatural," as that would be circular. Traditionally, philosophers have identified two features of mental phenomena – intentionality and consciousness – as essential (as that which makes them mental). Intentionality is that feature of mental states by virtue of which they have what we call a "content": what they are about. So, when I think that the peaches on my tree are now ripe, say, my thought – a state of my mind – is *about* the peaches on a particular tree. By virtue of having a content, thoughts can be true or false; in this case, true if those peaches are ripe, and false otherwise. Similarly, my desire for a ripe peach to eat has a content – that I eat a ripe peach – and, while not either true or false, is fulfilled or not depending on whether I do indeed eat a ripe peach. "Aboutness," then, is the core of intentionality.

One might have thought, both from the colloquial meaning of "intentional," as in "she did it intentionally," and also from the ordinary distinction between the natural and the artificial, that what is crucial to the contrast with the natural is the idea of purpose, or goal. That is, a natural object is one that has arisen from a process that had no inherent purpose or goal, whereas an artifact is something that was explicitly designed for a purpose. Teleology, serving a goal, might seem more to the point than intentionality, as philosophers conceive it.

However, there are good reasons to focus instead on intentionality. The main one is that teleology presumes intentionality, at least when it comes to minds; intentionality, then, is more fundamental than teleology. After all, what is involved in someone's building something for a purpose? Aside from the actual physical activity of building it, what makes it a purposive activity are the beliefs, desires, and intentions of the artisan. A hammer is designed for a purpose to the extent that someone represented to herself that a certain kind of device was needed and then envisaged – that is, represented to herself – that something shaped in such a way and made of such a material would satisfy the job description. So the ability to represent is prior, in this sense, to the ability to plan and execute the satisfaction of goals.

1 For an example of this use of the term "naturalizing," see Dretske (1995).

What's more, appeal to representation solves a kind of puzzle associated with teleology. It's common to think that the purpose for which an object is created counts as a kind of cause of its existence (the ancients refer to this as a "final cause"). But how can a nonexistent future state of affairs cause anything in the present? How do the future functions the hammer is going to fulfill explain how it came into existence? The answer to this puzzle is that goals and purposes play the causal role they do by virtue of their being represented in the minds of the producers. While the actual functions themselves are fulfilled after the object appears – how else could it be? – the representations of the functions occur prior to the appearance of the object. Hence, there is no puzzle after all.

The second feature of mentality is consciousness. By "consciousness," I mean one's subjective experiences, the sort of thing it is apt to describe in terms of "what it's like" for the subject having the experience.[2] Sensory experiences, such as the smell of a rose, the taste of chocolate, or the searing pain of a toothache, are "like something" for the subject having them. Most would argue that there is also "something it is like" for the subject to experience emotional states, like anger and love, and some even include episodic thoughts, such as consciously wondering whether one locked the door before leaving the house. For reasons to be presented later, this feature of mentality has proved resistant to explanations in terms of physical embodiment.[3]

To sum up, the core contrast with the natural, as I understand it, is with the mental. Of course, one may wonder, if the mental is contrasted with the natural, then is naturalism, as applied to the mental, just a contradiction in terms? Of course not, but this brings us to the next section.

Naturalism vs. Dualism

Materialists don't generally deny the existence of the mental.[4] Rather, their view is that mental phenomena, perfectly real though they be, are nothing over and above physical phenomena. As already mentioned, to make the doctrine precise requires elucidating both what is meant by "the physical" and what is meant by being "over and above." Both of these issues have spawned a great deal of literature, but I will provide a brief, and somewhat over-simplified, description of my own view on these matters.[5] I don't believe there is any good way to define "physical." The science of physics is the foundational study of the physical, and whatever it comes up with is ipso facto physical. The problem is that we don't know what physics will come up with in the future, so it's hard to see how we can get a genuinely substantive conception of the physical. In the end, is matter just a wave, a wave–particle combination, a field, or something we haven't yet considered? Fortunately, I think defining the physical isn't necessary for a proper characterization of materialism. When it comes to philosophy of mind, the doctrine of materialism can get by without a positive conception of the physical, so long as we have a positive conception of the mental; and this is what I characterized in the last section. Thus, the basic thesis of materialism can be stated as follows: The mental is nothing over and above the non-mental.

2 Thomas Nagel (1974) made this the standard way to express this aspect of mental life.

3 For debate over so-called "cognitive phenomenology," see Bayne and Montague (2011).

4 Though some do go that far: the so-called "eliminativists." See Churchland (1981), for example.

5 For an overview of that literature, and an extended defense of my own position on it, see Chapter 1 of Levine (2001).

Let's turn next to the phrase "over and above." What is it for one phenomenon to be something "over and above" another? I won't provide a precise characterization here, but rather give the general idea and flesh it out with some examples. Perhaps the basic idea is that of metaphysical construction. Some objects and states of objects are "constructible" from more basic objects and states. So, for instance, statues are constructible from materials such as clay and marble, which are then shaped according to the intentions of the sculptor. You put the material into the right shape (with the right intention) and you have a statue. In this sense, a statue is nothing over and above the relevant material being (intentionally) shaped the right way.

Another example, much closer to the case of the mental, is the relation between software and hardware in a computer. Programs are lists of instructions, and, given that instructions involve representational content (they involve telling something what to do), they embody a form of intentionality. Also, there is an abstract computational structure to the procedures determined by the program that is conceptually independent of the physical description of the computer's hardware. The very same program can run on computers constructed in a variety of different ways. Yet, the program is nothing over and above the physical processes of the computer. That is, once we determine which physical processes are taking place inside the computer, this determines which programs it is running. Put another way, we can explain how it is the computer carries out its programming function by appeal to the electronic hardware. That is, we can see how it is that the software is constructed out of the hardware.

Construction, in the way I'm using it here, has both a metaphysical and an epistemological side to it. On the metaphysical side is the fact that once the sequence of physical states of the computer is determined, its program is thereby determined as well. On the epistemological side, we can explain the ability of the computer to carry out its program by appeal to the physical states. So explanation and construction, or determination, go hand in hand here. When one phenomenon is constructible in this way from another, we will say that the former is nothing over and above the latter. Materialism (as it applies to the mind) is the doctrine that mental phenomena are constructible in this way from non-mental phenomena. In the case of human minds, that would mean that our mental lives are constructible from the electrochemical processes of our brains.

While no one denies that, in our sense, the software of a computer is nothing over and above its hardware, whether our minds are nothing over and above our brains is a matter of vigorous controversy. Materialists point to the computer analogy and claim that the relation of our minds to our brains is relevantly similar to that between a computer's software and hardware. Dualists, however, point to two crucial differences. First, the intentionality of computer states – their constituting representations – is derived from that of their human designers. That is, if one asks by virtue of what facts it is that a certain state of the computer represents what it does, it is reasonable to answer that the relevant facts include the intentions of the users and designers of the computer. But, of course, materialists can't claim that human intentionality is similarly derived, so they need another element in their story about how the mental is constructed out of the non-mental. Given that the non-mental items out of which mentality is supposedly constructed – the relevant biochemical ingredients – are not themselves invested with intentionality (they don't represent anything, aren't about anything), where does it come from?

Second, while this is a matter of some dispute, many philosophers, along with most people, are reluctant to attribute consciousness – subjective experience – to computers.

Few would claim that there is something it is like to be my Mac laptop.[6] So again, if there is nothing it's like to be a neuron, how can it be that there is something it's like to be a hundred billion of them? Answering these two challenges – how to construct both intentionality and consciousness out of materials that have neither feature – is the task facing the materialist.

Materialists do have a lot to say with respect to both challenges. When one thinks about what tools for constructing mentality are provided by non-mental objects and processes, the main item that emerges is this: some physical systems, by virtue of the causal relations they sustain internally among their states and externally with the environment around them, constitute "information processors." Materialists then try to show that, with enough sophisticated twists, a theory of the construction of the fully mental out of the non-mental is possible in terms of the causal relations that sustain information-processing.[7] This is a tall order, and there is no consensus on any particular version among materialists, much less agreement on the project's prospects for success. Suffice it to say that materialists believe some such theory can work and dualists deny this.

Let's return now to the question of naturalism. As just described, the materialist project is to show how it is possible to construct the mental out of non-mental ingredients. This project is often called the project of "naturalizing" the mind. Why "naturalizing," and not just "materializing" or "physicalizing"? The controlling idea here seems to be that mentality – intentionality and consciousness – is not to be found among the basic, fundamental ingredients of nature. By "basic" or "fundamental," I mean an ingredient that is not constructible from any of the others. So, if one holds this position with regard to the basic ingredients of nature, one will also hold that naturalism entails materialism. Naturalism, on this construal, is the claim that all phenomena are constructible from a certain restricted class of items; in particular, they are all non-mental items (probably non-normative as well, but this is a different issue). Dualism, which posits the mental as basic – the mental is not constructible from the non-mental – would then be opposed to naturalism.

I think this is one very reasonable way to construe naturalism, as it makes sense out of the "naturalizing" projects of materialists. Dualists, on this view, are non-naturalists, because they include "non-natural" items, like minds, among the basic constituents of the world. However, I do not think this construal of naturalism is appropriate in all contexts. For one thing, there are dualists who consider themselves naturalists. Obviously, in the sense we've attached to the term "naturalism" in this section, that would be a straightforward contradiction. Second, naturalism is often invoked in arguments in the philosophy of religion, where it is usually opposed to (at least traditional forms of) theism. As I will argue in the next section, for this purpose we need a different construal of naturalism, according to which dualists can indeed be naturalists.

Naturalism vs. Theism

I claimed earlier that the core idea of the contrast between the natural and the unnatural had to do with the contrast between the mental and the non-mental. One construal of this contrast is that naturalism is the doctrine that the mental can be constructed out of the

6 What about Data, from Star Trek? Yes, sufficiently complex and human-like machines might qualify for most people, but how would we tell? This is a complicated issue, to be sure. See the discussion of the question of how we would tell in Levine (2001) and Block (2002).

7 Dretske (1995) is a good example of an attempt along these lines.

non-mental. So, while naturalism does not deny the existence of the mental, it denies that the mental appears among the fundamental or basic ingredients of reality or nature. Dualism, by contrast, insists that the mental is found among the fundamental ingredients. Let's call this notion of naturalism, according to which it is opposed to dualism, "narrow naturalism."

However, there is another kind of contrast that one might have in mind between the natural and the unnatural, and this too involves the distinction between the mental and non-mental, but in a different way. One question regarding the mental is whether it is included among the fundamental entities out of which all nature is constructed. Another is whether the mental in some way serves as the foundation for everything else, particularly the non-mental. Let me explain what I mean here by explicitly relating it to the doctrine of theism.

On the traditional Judeo-Christian conception of God, God is an immaterial entity who has thoughts and plans. God created the world, with everything in it, for the purpose of realizing His plan (whatever that is). On this view, then, God's thoughts exhibit intentionality – they are about something; for instance, the things He intends to create and His plans for them – and they involve teleology as well, as it is His purposes that determine what there is and what laws govern it. On this conception of the world, then, intentionality – both in the philosophical sense of "aboutness," or representational content, and in the ordinary sense of doing something for a purpose – underlies everything there is, including the laws that govern them. Mind, and intentionality, constitute the core of reality, on this view.

It is this sort of view, which I'm calling "theism," that I want to contrast with "broad naturalism," which is to be distinguished from the narrow naturalism characterized in the last section.[8] What I said about mind constituting the core of reality suggests the doctrine of idealism, a view traditionally associated with Bishop Berkeley.[9] On Berkeley's idealism, material things are really mental entities in disguise, ideas contained within minds. Everything is mental on this view, and the order maintained within the realm of ideas that gives the appearance of an external, non-mental reality is sustained by God, who encompasses all ideas within Himself.

While Berkeley's idealism is certainly a version of the non-naturalism/theism I have in mind, it isn't the only version. Most theists, even of the very traditional sort, are not idealists. They believe that among God's creations is inanimate matter, out of which all physical objects are made. Their theism is opposed to (broad) naturalism[10] because they believe that the source of matter (as well as of minds other than God's), along with the laws governing its behavior, is God's intentional plan. In this sense, nature is at bottom intentional, not because all entities in it are intentional, or mental, but because everything that exists is explained ultimately by appeal to intentions (i.e., God's intentions).

Naturalism is the view that all of nature is determined by laws and processes that have no plan, no intentions, at their foundation. So, for instance, if you want to know why unsupported objects fall to the earth, you appeal to gravity. This, both naturalists and theists can do. However, suppose you ask, "Why is there gravity?" Or, "Why does gravity work precisely the way it does?" While physics may have answers to these questions, at some point, if you keep pushing for explanations, physics hits rock bottom: that is, a set of fundamental laws, and

8 While theism is the most obvious and prominent form of non-naturalism, and the one with which I'll be concerned in this paper, note that there can be nontheistic versions of non-naturalism as well.

9 See Berkeley (1713/1979).

10 For the rest of the chapter, I will drop the term "broad" and just use "naturalism" for this doctrine, including the qualifier "broad" only when context requires it.

possibly some singular events (such as the Big Bang), which are treated as just brute and given. The answer to why things are like this is just: because that's how the world is constructed. Whereas the naturalist bottoms out in her series of natural explanations with brute, and thus inexplicable, laws and facts – the very fundamental, or basic, laws and facts – the theist explains the entire system, including the laws and facts treated as fundamental within physics, by appeal to God's intentions. In this sense, then, the debate over naturalism is a debate about whether nature is an explicitly designed system or whether it is a given, brute fact. For the naturalist, then, nature may be intelligible to the extent that we can someday figure out what are its most fundamental laws and facts, but it is ultimately unintelligible in the sense that there won't be any overriding plan – intentionally conceived – that explains why it is precisely the way it is. To sum up, naturalism is the view that brute fact – just the way things are – constitutes the foundation of nature, where its denial, most prominently theism, sees a deep intelligibility and explicit design at this foundation.

Let's turn to dualism and see on which side of this naturalism–theism divide it falls. My claim is that one can find dualists on both sides. Clearly, theists, at least of the traditional variety, will be dualists (if not idealists). God's mind, at least, and likely human minds as well, will not, on this view, be constructed from non-mental ingredients. God, after all, is allegedly immaterial, so there must be mental facts – those involving God's ideas and plans – that are indeed something over and above the non-mental. Also, on many theistic doctrines, humans possess an eternal soul that is fundamentally mental, and not constructed from non-mental, physical elements. So it's easy to see how dualists fit naturally into the non-naturalist, theistic camp.

However, not only is it the case that dualists need not be theists, but they can be naturalists as well. Indeed, David Chalmers (1996) calls his form of dualism, "naturalistic dualism." I also consider myself a naturalist but also a dualist (or, at least, an antimaterialist). Let me describe the position that I espouse, and why I consider it to be a version of naturalism, at least on this second interpretation of it.

Naturalistic dualism is the position that the mental is something over and above the non-mental; in fact, mental phenomena are among the fundamental constituents of reality. There are a lot of different versions of this position. Some hold that minds are immaterial entities – very much like the souls of religious doctrines – that are attached in some way to bodies and interact with them. Others hold that there are fundamental mental properties that are possessed by, or instantiated in, physical objects. The former view is usually called "substance dualism," whereas the latter is called "property dualism." Among property dualists, some believe that the fundamental mental properties are instantiated in the very simplest physical objects (like elementary particles) – these are called "panpsychists" – while others hold that it is only certain highly complex physical systems like brains that instantiate mental properties – these are called "emergentists." As I consider myself an emergentist, let me describe that position briefly and explain why it counts as naturalistic.

Many of the capacities that we consider mental – such as our ability to speak language, pick up information concerning our immediate environment through our sense organs, and voluntarily move our bodies – seem to be explicable in terms of physically embodied information-processing systems. Oversimplifying greatly, we can explain these mental capacities in ways that are very similar to the ways we explain the capacities of computers to solve problems that seem to require intelligence. This much is consistent with materialism, as the capacities in question are shown to be constructible from non-mental entities: neurons and their basic chemicals.

However, when it comes to conscious experience – such as seeing and smelling a rose, feeling a toothache, or daydreaming about one's upcoming vacation – it seems that the situation is quite different. If we ask, "What explains the fact that we have subjective experience?", it's not at all clear that appeal to the physical events underlying our sensory interaction with the world, together with the internal neural processing that follows, helps much. For instance, consider what it's like to see a red rose in sunlight, or feel the pain of a toothache. While certain physical conditions can be cited as the causes of such experiences – light bouncing off the rose into our eyes, or inflammation near the nerve of the tooth – what we are looking for are the physical states that ultimately constitute the experiences in question. Presumably, these lie deep within the brain. But why should there be something it's like to have these brain states? What about the brain states explains not only their processing information regarding the rose and the tooth, but also their making it the case that one *feels* the pain and actually *experiences* the red rose? It seems possible, for instance, that a computer programmed to detect red roses and inflamed teeth could do so without experiencing either redness or pain; that is, without there being anything it is like to be that computer.

Of course, materialist philosophers have responses to this argument, and the debate is long-standing and complex.[11] But my position is that the best hypothesis of why we cannot explain conscious experience in terms of the physically embodied information-processing of the nervous system and brain is that the phenomenon is "emergent." That is, rather than being constructible from the underlying physical processing, there is just a basic law of nature such that when certain complex configurations occur, conscious experience also occurs. Assuming that this psychophysical law – a law that relates certain physical phenomena to certain mental phenomena – is itself a fundamental or basic law of nature, then the mental phenomena themselves will count among the fundamental ingredients of reality. By including mental phenomena, like subjective conscious experiences, in the inventory of basic phenomena, we have abandoned materialism. We defined materialism (with respect to the mind) as the doctrine that all mental phenomena are constructible from non-mental phenomena (hence, mental phenomena are not basic in nature). Emergentism, which posits mental phenomena among the basic phenomena, is inconsistent with materialism.

So, is emergentism, as I've just described it, a form of naturalism? It is, so long as the psychophysical laws that relate physical phenomena to mental phenomena are brute, fundamental, and therefore inexplicable. To repeat, the core idea of naturalism is that reality bottoms out in brute facts and laws. They are brute because they are not explicable by any more basic laws – after all, we're talking about the most fundamental laws – and because they are not explained as the product of any intentional design. For emergentists, intentions certainly exist – and may even be among the fundamental entities – but they are themselves the outcome of nonintentional, therefore purely *natural* processes. So long as the ultimate laws introducing mentality into the world are not themselves the result of intentional design, whether they are treated as basic, as the emergentist claims, or as constructible from non-mental phenomena, as the materialist claims, their existence is consistent with naturalism.

11 Among the best materialist strategies, to my mind, is the "phenomenal concept strategy," so named by Stoljar (2005) and pioneered by Loar (1990). For discussions of this strategy, both pro and con, see Alter and Walter (2006).

To sum up this section, there is a sense of "naturalism," broad naturalism, that can include dualism. To be a (broad) naturalist is to view the entirety of nature as a system of entities and laws that itself arose "naturally"; that is, not as the product of intentional design. For the theist, nature is a kind of artifact, sharing with our artifacts the feature of having been produced by explicit, intentional design. Naturalism opposes this conception, providing room for intentionally designed artifacts within nature, but rejecting the view that nature itself is an artifact.

Naturalism Defended

In "Flavors, Colors, and God," Robert Adams (1987) presents an argument to the effect that if you reject narrow naturalism, you ought also to reject broad naturalism (though this is not the way he puts it). That is, he thinks that the standard arguments against materialism that appeal to the special, irreducible nature of conscious experience also support theism, and thus abandonment of the more general form of naturalism described at the start of this chapter. In this final section, I want to argue that maintaining broad naturalism contra theism is quite reasonable, even if one has abandoned narrow naturalism or materialism.

Adams argues as follows. Let's assume that when it comes to conscious experience, materialism is false. In particular, he contends that those qualitative properties of our sensory experiences – the flavors and colors mentioned in the title of his paper – are not identical to, or reducible to, any physical, non-mental properties. Let's accept without question this portion of his overall argument.

Even though experiential qualities like the taste of sugar or the way a ripe red tomato looks are not reducible to any non-mental properties, it is still clear that there is a tight connection between these experiences and the electrochemical states of our nervous systems. If someone is visually experiencing red, you can (or, at least, vision scientists can) predict quite a bit about the state of that person's visual cortex. If you were to remove the relevant part of the visual cortex, color vision would go away. So, while experienced qualities are not identical to, or reducible to, the neurophysiological properties of one's brain, there is a robust and lawful correlation between the two kinds of property. No brain, no experience. If there's experience, then there's some corresponding brain activity.

Materialists and dualists agree that there are robust correlations among physical (i.e., non-mental) states and conscious experiences. However, the materialist believes that we can explain these lawful correlations in more basic, non-mental terms. The dualist, on the other hand, believes that these correlations reflect basic, brute facts; they are included among the fundamental laws of nature.

With respect to Adams's argument (and also according to my own considered opinion), we are siding with the dualist here. Thus, we accept that the robust correlations that hold between physical states and types of experience are, in a sense, brute and inexplicable. There is no more basic level of analysis to which we can appeal – one that descends beneath the level of the mental to deal only in the non-mental – in terms of which the tight connection between some particular type of neural process and the experience of seeing red, say, can be explained. It's just a basic fact that the former causes the latter, or that the latter emerges from the former.

This is where Adams's argument enters the picture. Adams claims that it's highly implausible that there is no explanation of the tight connection between physical states of the brain

(or nervous system) and types of experience. Surely, he argues, there must be some reason that this particular kind of neural activity gives rise to experiences of red and that other kind of neural activity gives rise to experiences of green. But if we reject – as we're assuming now – that this correlation can be explained in terms of more basic non-mental properties, then what's left? Well, what about God? Why not say, as many early modern philosophers (like Descartes and Locke) did, that what explains the correlations between physical state types and experience types is God's decision to pair just those items, and not different ones? So, for instance, if you want to know why neural activity of this sort causes reddish experiences and neural activity of that sort causes greenish experiences, the answer is that God – for His own reasons – decided to connect them in that way. Nothing forced that upon Him, and He could have done the reverse. The argument, then, is that by positing God as the creator of everything, we can explain these apparently brute correlations as results of His intentional action, in the same way that all of nature is explained as the result of His intentional action. Thus, dualism gives us reason to adopt theism and reject naturalism of any sort.

I do not find this argument convincing. Adams seems to assume that an explanation of the correlation between type of neural activity and type of conscious experience is necessary, and that in the absence of a materialist explanation, the most plausible hypothesis is the theistic one. However, he ignores the possibility that this correlation just doesn't have an explanation; that is, that it is a fundamental or basic law of nature. Notice that the most fundamental laws of physics are not amenable to explanation: after all, in terms of what could one explain them, if they are indeed the most fundamental laws already? Nature, then, contains brute facts and laws. So why couldn't the lawful correlations between types of neural activity and types of conscious experience be among them?

I can imagine two replies Adams might make. First, he might say that theism not only explains the otherwise brute correlation between physical states and mental states, but also provides an explanation for those supposedly fundamental and inexplicable laws of physics themselves. Everything that exists, and all the laws that govern it, is the result of God's plan, His intentional decisions and actions. So theism provides explanations that naturalism fails to provide.

This reply begs the question against naturalism in a very fundamental way. As I characterized the contrast between naturalism and theism in the last section, it comes down to whether or not one must find reason – in the sense of an intentionally characterized purpose or plan – underlying nature or whether nature in some sense "just is." The naturalist says that there is a fundamental level of reality that is not a product of reason, and claims that on this basis all other levels of reality can be explained in terms of it. The theist, for her part, insists that there must be a reason – an intelligible and intentionally entertained purpose – why things are ultimately the way they are. So to defend theism by appealing to this position is to beg the question against the naturalist. The naturalist denies that explanations must ultimately bottom out in reasons (of the intentional variety).

What's more, even the theist has to grant a certain level of brute and inexplicable reality. After all, what is the explanation for God and His purposes and designs? Aren't they fundamental? It doesn't seem that one can appeal to any reasons that God exists or has the plans He has, unless one appeals to yet further deities, and that way lies both infinite regress and polytheism. Theists generally reply that since God is a necessary being – God exists necessarily, unlike all other entities in space and time, which exist only contingently – God's existence requires no explanation. But I find this evasive. First, it's not at all clear what a necessary being is supposed to be, and unless one can make sense out of this, one can't

appeal to it to explain other things. Second, given that God's existence isn't conceptually necessary – it's coherent to imagine that there is no God – then we can still ask why there should be a necessary being of this sort. That is, we cut off explanation at God's existence only by stipulation.

The other reply Adams might make is more plausible. The problem with treating the relevant physical–mental correlations as basic is not that all basic laws need an explanation but that correlations that obtain at such a high level of composition – at the level of brains, not elementary particles and fields – are especially difficult to accept. Indeed, this appeal to the oddity of having brute and basic correlations at this level of aggregation in the world is one of the materialist's arguments against the dualist. There must be an explanation in non-mental terms, the materialist argues, because we expect the only fundamental laws to obtain at the most fundamental, or simple, level of aggregation.

I am sympathetic to this consideration, and indeed find it one of the strongest considerations in favor of materialism, though I think there are even stronger reasons for rejecting materialism. So why not go all the way and abandon all forms of naturalism? This strikes me as rash. Yes, there is something odd about there being fundamental laws of nature at the level of the brain. But I find it odder to embrace the entire theological account of the origins of nature as a whole, to see God's intentional design behind everything, just to explain this one admitted oddity about conscious experience. Perhaps the best one can say – and this might be agreeable to Adams himself – is that if one has other compelling reasons to reject naturalism and adopt theism, then the ability to explain the correlations between types of neural activity and types of conscious experience, as the unfathomable choices of an intentional agent whose reasons are beyond our comprehension, adds a little bit of support. But not finding any of these other reasons compelling, I find the postulation of fundamental laws relating brains to consciousness the simpler and more conservative hypothesis by far.

References

Adams, R.M. (1987). *The Virtue of Faith and Other Essays in Philosophical Theology*. Oxford: Oxford University Press.

Alter, T. and Walter, S. (2006). *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.

Bayne, T. and Montague, M. (2011). *Cognitive Phenomenology*. Oxford: Oxford University Press.

Berkeley, G. (1713/1979). *Three Dialogues between Hylas and Philonous*, edited by R.M. Adams. Indianapolis: Hackett.

Block, N. (2002). The Harder Problem of Consciousness. *Journal of Philosophy* XCIX: 391–425.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy* LXXVIII: 67–90.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: Bradford Books/ MIT Press.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.

Loar, B. (1990). Phenomenal States. In *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, edited by J. Tomberlin, pp. 81–108. Atascadero: Ridgeview.

Nagel, T. (1974). What Is it Like to Be a Bat? *Philosophical Review* 82: 435–450.

Stoljar, D. (2005). Physicalism and Phenomenal Concepts. *Mind and Language* 20: 469–494.

16

Epistemological Naturalisms

C.S.I. JENKINS

One traditionally prominent kind of epistemological naturalism centers on the rejection of *a priori knowledge*. Its motivating idea is that the very idea of *a priori* knowledge is unscientific, because science is a purely empirical enterprise. Since anything that is unscientific is thereby unacceptable to a naturalist, the thought continues, epistemological naturalists must reject the *a priori*. Something like this line of thought has also been applied to *epistemic normativity* and *armchair knowledge*, both of which have been treated as similarly inimical to epistemological naturalism.

In this chapter, I will argue that the real picture is considerably subtler than this caricature suggests. The chapter begins with a brief classificatory sketch of different kinds of naturalism, intended to help give the reader a sense of how to navigate the vast range of naturalisms (or at least, uses of the word "naturalism") in contemporary philosophy. I will then focus in on some of the forms of naturalism that have been especially significant in connection with epistemology in particular. Finally, I single out one kind of epistemological naturalism for further examination, and survey a number ways in which, despite being a substantive commitment, naturalism in this sense is considerably less radical than epistemological naturalism is often taken to be. I argue in particular that this option, which I call *lightweight epistemological naturalism*, is consistent with robust (nondeflated, noninstrumental) epistemic normativity, consistent with *a priori* knowledge, and noncommittal with respect to the prospects of negative experimental philosophy's critique of armchair knowledge.

It is not part of my goal in this chapter to convince readers that any version of epistemological naturalism is correct. The aim is, rather, to show that there is no clearly identifiable single thesis best deserving of the name "epistemological naturalism," and that getting clear about the various distinct *threads* in the fabric of traditional epistemological naturalism makes lightweight epistemological naturalism an available option.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

Classifying Naturalisms

The array of philosophical views – and other things – that go by the name of "naturalism" is huge, and ever-expanding. To help orient the reader and locate the main issues to be addressed later in this chapter, I will sketch here my preferred way of approaching the classification of naturalisms. This sketch is intended to illuminate the contours of the range of options, but is not comprehensive in the sense of describing all the actual and possible positions within this structure.[1]

It is relatively common to see a distinction made between *ontological* naturalism and *methodological* naturalism (see, e.g., Papineau 2007). A first-pass characterization of these two naturalisms will suffice for current purposes. Roughly, ontological naturalism says that only natural things exist, where "natural things" means the kinds of thing characteristically posited by the natural sciences (contrasted not with "artificial things," but with supernatural or otherwise "spooky" things). Methodological naturalism, on the other hand, says that philosophical inquiry is to be conducted in the manner of natural science (though what exactly this amounts to is a subject for further discussion; we shall return to it later). Ontological and methodological naturalism make for a comfortable pairing, though each could be consistently held without the other.

These are certainly two of the most significant forms of naturalism. But there is a lot more variety out there than this, and my classificatory sketch aims at equipping philosophers with the conceptual resources needed to make much more fine-grained distinctions among the naturalisms.

Most (perhaps all) of the full range of naturalisms ultimately have some sort of connection with science, or at least with someone's idea of science. But beyond this suggestion, which is intentionally *very* vague, features shared by all or even most forms of naturalism are hard to find. Insofar as it is possible to draw a distinction between "natural" and "social" sciences, there is (as the nomenclature might suggest) sometimes an association between naturalism and the *natural* sciences in particular. But philosophical naturalism has also quite often been associated with science more generally, including the social sciences.

While "naturalism" often refers to a doctrine or thesis, it does not always do so. Sometimes, for example, "naturalism" refers to a stance, an approach, a methodology, or an injunction (e.g., an injunction to use a certain kind of methodology). Sometimes naturalism is characterized by an *attitude*, such as an attitude of "respect for the findings of science, and for the progress that has been made by science" (Nolan 2005, 10). In other senses of "naturalism," mere *failures* to do something (e.g., failing to believe in non-natural things) may be sufficient for being a naturalist.

Doctrine naturalisms can be divided into the first-order philosophical and the metaphilosophical. Metaphilosophical doctrine naturalisms are theses about philosophy; for example, the claim that philosophical inquiry is part of the scientific enterprise. First-order doctrine naturalisms are philosophical theses, but are not theses about philosophy. They may *apply* to philosophy, among other things, but they are not about philosophy in particular. For example, the claim that there is no *a priori* knowledge is a first-order doctrine naturalism.

First-order philosophical naturalisms include quite general views in ontology, such as the view that only natural things exist ("ontological naturalism" in the familiar bipartite

1 This classificatory sketch is developed a little more in Jenkins (2013), for those who wish for further details.

distinction). There are also more subdiscipline-specific ontological naturalisms. Some of these are still ontological, such as the view that "the ontology of mathematics is the mathematical ontology of our best natural science" (Paseau 2008, § 6.1). Other views described as "naturalism" are first-order and *not* ontological; for example, that "moral words have the same *meaning* as certain combinations of non-moral expressions" (Hare 1995, 340), that "all facts are natural facts" (Lenman 2006, § 1), or that "science, properly so-called, cannot involve religious belief or commitment" (Plantinga 1996, 178). There are many other first-order naturalisms too, scattered across subdisciplinary areas. The fact that not all first-order naturalisms are ontological is one of the main reasons I do not think philosophers should rest content with the simple ontological/methodological distinction among naturalisms.

The range of extant first-order *epistemological* naturalisms includes such varied theses as the claim that "[i]t is rational to believe (only) our best scientific theories" (Colyvan 2009, § 4), the claim that "all epistemic facts are natural facts" (Feldman 2001, § 4), and the claim that there is no *a priori* knowledge (Devitt 2005). These are better classified as epistemological than *methodological*, since they directly concern epistemological notions like knowledge, justification, and rationality, rather than method of inquiry.[2] This is another key reason why we need more detailed categories than are provided by the ontological/methodological distinction.

Metaphilosophical naturalisms can be subdivided into the descriptive and the normative. Descriptive metaphilosophical naturalisms include the view that philosophy is literally and already part of the scientific enterprise. Normative metaphilosophical naturalisms include the view that philosophy *should* be part of the scientific enterprise, whether or not it already is (this is the "methodological naturalism" of the familiar bipartite classification).

Figure 16.1 illustrates the classificatory schema outlined in this section. Many self-identified naturalists accept bundles of naturalisms, scattered across different nodes of the chart. It is nevertheless important to separate distinct elements in these bundles so that the relationships between them can be made clear.

Some Epistemological Naturalists and their Naturalisms

Many of the naturalisms that have been of significance for recent epistemology are clustered in the bottom left-hand node of Figure 16.1. These are (subdiscipline-specific) epistemological first-order doctrine naturalisms. However, as I shall illustrate in this section, there are also some epistemologically significant naturalisms that belong in other nodes. These have tended to be bundled together, into naturalistic package deals, by most of the prominent self-identified epistemological naturalists.

As with the previous section, I won't aspire to a comprehensive presentation of extant epistemological naturalisms here. My strategy will be to present and discuss a sample of prominent recent epistemological naturalists, and the clusters of views they espouse under the "naturalist" banner.

I'll start with Quine (1969), and his argument that epistemology should be "naturalized." By this he meant that epistemology traditionally construed (as the analytic investigation of

2 Of course, there are *also* first-order methodological naturalisms, such as the view that "the mode of inquiry typical of the physical sciences will provide theoretical understanding of the world, to the extent that this sort of understanding can be achieved" (Stoljar 2009, § 12).

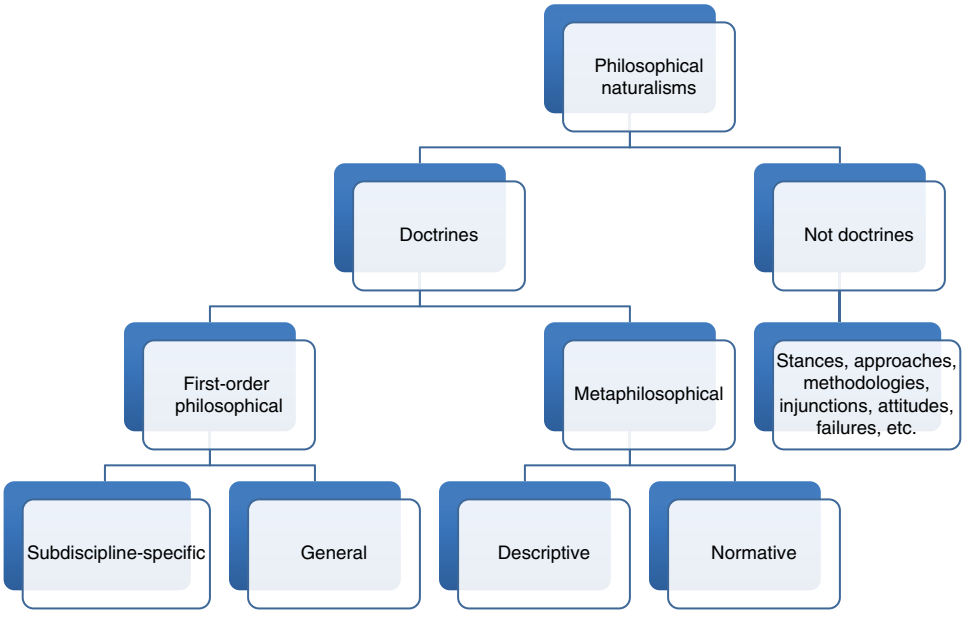


Figure 16.1 Classification of different naturalisms.

knowledge, justification, etc.) should be *replaced* with empirical psychology: the investigation of how we (as a matter of fact) respond to sensory stimuli in the formation of our beliefs. And like all inquiry, Quine thinks that psychology should be conducted using only the methods of natural science, which according to him are wholly empirical and include nothing *a priori*. (It might be more standard now to think of what we call "psychology" as a social science. Quine's conception of psychology seems to have tended toward what might now be labelled "cognitive science.")

This radical position enabled Quine to accommodate at least three different motivating naturalistic views concerning epistemology, to which he was evidently committed (in Quine 1969 and elsewhere in his work):

(1) A metaphilosophical, normative, methodological naturalism, which says that the methods of the natural sciences are the only correct methods by which to pursue epistemological inquiry.
(2) A first-order metaphysical naturalism, which holds that there are no normative facts or properties such as are apparently studied by traditional epistemology.
(3) A first-order epistemological naturalism, according to which there is no *a priori* knowledge of the kind that is apparently required in order for the traditional analytic investigation of knowledge to be viable.

The first two of these make a fairly natural pairing: (2) can be supported by (1) in combination with the thought that the natural sciences do not study normative subject matters. The third is famously defended in Quine (1951). At first blush, it might seem awkward to maintain an epistemological view like (3) in conjunction with the *rejection* of epistemology as a respectably area of study, but given that (2) is a purely negative claim, it can be understood as analogous to the claim that there is no phlogiston, which one can comfortably maintain

alongside the rejection of phlogiston theory as a respectable area of study. It is more problematic for Quine to maintain that there is any empirical knowledge or justification in the light of his commitment to (2).

In later work, Quine defends a different bundle of naturalistic views about epistemology. Instead of flat-footedly arguing that we should replace (normative) epistemology with (descriptive) psychology, he tries to render epistemic normativity naturalistically respectable by *reducing* it to the descriptive. So (2) is replaced with the more moderate:

(2*) A first-order metaphysical naturalism, which holds that the normative facts or properties studied by traditional epistemology are reducible to instrumental facts (which can be investigated using empirical scientific methods).

"Instrumental" facts in the intended sense are facts about what can be used as a means to what ends. Quine says that normative epistemology is "the technology of truth-seeking…a matter of efficacy for an ulterior end, truth" (1986, 665). And thus, "the normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed" (1986, 664–665), the relevant "terminal parameter" being the goal or end of finding out the truth. The descriptive facts with which Quine thinks epistemically normative claims are to be associated are instrumental facts about what is conducive to the end of discovering the truth.[3] Quine appears to maintain his commitment to (1) and (3) over the course of the move from (2) to (2*).

Quine's two bundles of epistemological naturalisms, namely (1), (2), (3) and (1), (2*), (3), are classic. But they are by no means the only ways that naturalistically minded philosophers have addressed epistemology. The next bundle of naturalisms I'll consider here is that of Maddy, who describes her naturalism as "a descendant of Quine's" (Maddy 2000, 92), and who, like Quine, applies it to first-order epistemology as well as to inquiry more generally.

Maddy sometimes characterizes her naturalism as "an approach" rather than "a theory" (see, e.g., 2000, 108). The naturalistic approach she favors is that of the "Neurathian sailor, working within science to understand, clarify and improve science," who will "treat philosophical questions on a par with other questions, as far as this is possible" and "respond with befuddlement" if asked to do otherwise (2000, 108). So at least one of the naturalisms to which Maddy is committed belongs on the nondoctrinal side of Figure 16.1.

But Maddy's naturalism also evidently involves commitment to some theses; for example, as part of her bundle of naturalistic commitments, Maddy maintains that there is no "first philosophy" of the kind Descartes believed in, and that any "two-level" approach to inquiry on the model of Carnap's is a mistake.

To fill in a little background, the Cartesian conception of "first philosophy" is a conception on which the core questions of philosophy (including, centrally, epistemological questions) are to be pursued using the rational faculties and are importantly *foundational*, in the sense that they are epistemically and methodologically prior to all other inquiry.

The Carnapian position that Maddy rejects is roughly that certain questions (those amenable to *a priori* philosophical analysis) are settled entirely by the choice of a linguistic framework (approximately: the adoption of a language in which to conduct inquiry), and that linguistic frameworks are chosen on the basis of pragmatic considerations. On this view, there are two kinds (or "levels") of inquiry: the *a priori* philosophical investigation of a linguistic framework and the empirical, scientific investigation of questions that are

3 See Kelly (2003) for discussion and critique of instrumentalist accounts of epistemic normativity.

formulated using the *language* of some framework but not *settled* by the framework alone. For example, Carnapians often take it that arithmetical issues (such as whether 2 + 2 = 4) are settled *a priori* through the adoption of a linguistic framework suitable for talking in numerical terms, while questions about the state of the physical world (such as how tall Everest is) are not settled by these means but through empirical investigation.

In contrast to these rejected positions, Maddy holds that all inquiry is in fact part of the empirical scientific enterprise. By Maddy's lights, any attempt to construe philosophy "first," or treat it as a different kind or "level" of inquiry, is a mistake that is appropriately met with "befuddlement."

Maddy also takes her naturalism to have first-order epistemological components (or, at least, direct consequences); in particular, she thinks that it leaves no room for any *a priori* knowledge. For Maddy, the *a priori* is bound up with the postulation of a separate "level" of inquiry different from the empirical, scientific kind. She urges that we should adopt "a one-level view that fails to underwrite the a priori–a posteriori distinction" (2000, 114), adding that though "naturalism may doom the time-honored notion of a priori knowledge," she is "quite willing to accept" this cost (2000, 114).[4]

To summarize, we can think of Maddy's bundle of epistemological naturalisms as including at least the following:

(1) *From Quine's list:* A metaphilosophical, normative, methodological naturalism, which says that the methods of the natural sciences are the only correct methods by which to pursue epistemological inquiry.
(3) *From Quine's list:* A first-order epistemological naturalism, according to which there is no *a priori* knowledge of the kind that is apparently required in order for the traditional analytic investigation of knowledge to be viable.
(4) A nondoctrinal naturalism, which adopts the approach, when doing epistemology, of working within science to clarify and understand science.

Next let's consider Kornblith. Like Maddy and Quine, Kornblith rejects the idea that philosophy (epistemology included) should come "first" in the sense of "being more fundamental than." Kornblith argues that a big sweep of epistemology has made a serious mistake in focusing on the *a priori* investigation of the *concept* of knowledge, instead of the natural phenomenon that is knowledge itself (see, e.g., 1999, 161).

Among Kornblith's naturalistic commitments concerning epistemology are:

(5) A metaphilosophical, descriptive, methodological naturalism, which holds that epistemology is "very much a part of the scientific enterprise" (1999, 163).
(6) A metaphilosophical, normative, methodological naturalism, which holds that epistemologists should study knowledge (not the concept of knowledge), and should do it using empirical (not *a priori*) methods.

While Kornblith does not outright reject the *a priori* as Quine and Maddy do, he does in the course of discussing his reasons for (5) and (6) commit himself to some first-order epistemological claims concerning *a priori*. For example, he commits to:

4 Because Maddy is discussing the philosophy of mathematics in this work, this is a particularly surprising break with tradition. Mathematics is often considered a (or even *the*) core example of an *a priori* discipline.

(7) A first-order epistemological naturalism, according to which nothing substantive can be known *a priori* about natural kinds.

A *natural kind* in the sense Kornblith has in mind here is a grouping or category that is created by nature, independent of the artificial groupings or categories for which humans are responsible. For example, *oxygen* is usually taken to be a natural kind, but *opera* is not. According to Kornblith, we can't learn anything much about natural kinds through *a priori* conceptual inquiry. Instead, they must be approached by way of "straightforward empirical inquiry" (Kornblith 2007, 47).

Kornblith also thinks that knowledge is a natural kind, so we might add to Kornblith's bundle of epistemological naturalisms:

(8) The first-order metaphysical view that knowledge is a natural kind.[5]

I take (8) to be deserving of the label "epistemological naturalism" insofar as it is recognizably a claim to the effect that knowledge, while sometimes thought to be problematically normative and hence possibly "spooky," is in fact a part of the natural world.

However, Kornblith stresses that it is not actually essential, for his dialectical purposes in arguing for (6), that knowledge be regarded as a natural kind:

> It really doesn't matter…whether knowledge and other targets of philosophical analysis are natural kinds…[E]ven if the topics of philosophical interest typically correspond to…socially constructed kinds, it remains true that the concepts of the folk, and the concepts of philosophers as well, need not accurately characterize these socially constructed categories. Just as any individual's concept of aluminum may contain substantial errors or omissions, any individual's concept of a semiconductor, or Chippendale furniture, or of socially constructed categories generally, may contain substantial errors or omissions. So the gap between concept and category does not disappear simply because we have moved from natural kind concepts to socially constructed ones. And once we recognize that our concepts, whether the concepts at issue are those of the folk or of theoreticians, may fail to characterize the categories they are concepts of, the philosophical interest of our concepts thereby wanes. (2007, 36–37)

Kornblith's commitment to (6), then, is supposed to be independent of his commitment to the metaphysical thesis that knowledge is a natural kind.

To conclude this section, I'll discuss one more bundle of naturalistic views, championed by some pioneers of the movement known as "experimental philosophy." While not all experimental philosophers apply the label "naturalist" to themselves, Goldman (2013) rightly points out that in various respects experimental philosophy is appropriately identified as a subcategory of naturalism.

In their classic 2001 paper, experimental philosophers Weinberg, Nichols, and Stich offer resistance to what they take to be a (or perhaps the) traditional method for pursuing epistemological (and other philosophical) questions, namely reliance on *intuitions*. Exactly what they are resisting here is a matter in need of careful discussion, though it cannot be addressed in detail in this chapter.[6] What Weinberg, Nichols, and Stich explicitly *say* is that an epistemic intuition in the sense they intend is a "spontaneous judgment about the epistemic properties of

5 Note that, though this is a metaphysical claim, because it concerns the *metaphysics* of knowledge, it is epistemological too, because it concerns the metaphysics of *knowledge*. The subdisciplines of philosophy are tightly interconnected and overlap in many places.

6 See, e.g., Cappelen (2012) for extended discussion of some of the issues.

some specific case" (2001, 432). For them, having intuitions is an individual activity, one that can be conducted from a philosopher's armchair without asking anyone else what they think.

By contrast, the methodology they employ in their paper involves gathering empirical data about the judgments of multiple people through the use of surveys and then analyzing these data for statistically significant patterns. One of the principal conclusions of Weinberg, Nichols, and Stich (2001) is that by these experimental means they have discovered results which cast serious doubt on the viability of the traditional method. They write:

> [A] sizeable group of epistemological projects – a group which includes much of what has been done in epistemology in the analytic tradition – would be seriously undermined if one or more of a cluster of empirical hypotheses about epistemic intuitions turns out to be true…[W]hile the jury is still out, there is now a substantial body of evidence suggesting that some of those empirical hypotheses *are* true. (2001, 429)

Most notably, they claim to have discovered evidence suggestive of variation in intuitive judgments between different ethnic groups. This, they think, suggests that the traditional method of relying solely on one's *own* intuitions is unduly parochial. This sort of approach has been classified as "negative experimental philosophy" (in contrast with forms of experimental philosophy which do not criticize "traditional" methods but merely claim that experimental methods may have something distinctive to *add* to the philosophical enterprise[7]).

To categorize the views of Weinberg, Nichols, and Stich (2001) in terms of the classificatory system of this chapter, they espouse a bundle of naturalisms that includes the following:

(9) A metaphilosophical, normative, methodological naturalism, which says that the traditional method of reliance on armchair intuitions is an incorrect method by which to pursue epistemological (and, more broadly, philosophical) inquiry.
(10) A first-order epistemological naturalism, according to which armchair reliance on intuitions is unreliable, and hence not conducive to knowledge.
(11) A nondoctrinal naturalism, which adopts the approach of pursuing epistemological inquiry by means of surveys and statistical analysis, instead of reliance on one's own armchair intuitions.

All of these are variants on, but substantively different from, forms of naturalism we have seen already. One especially noteworthy point of departure from some of the related forms of naturalism covered earlier in this section is that armchair intuitions are not assumed by Weinberg, Nichols, and Stich to be *a priori*. They are spontaneous judgments made when presented with a description of a case, but, for all they say, such judgments may or may not be based on previously acquired empirical beliefs and/or evidence.

Lightweight Epistemological Naturalism I: Norms and the *A Priori*

In earlier work (e.g., Jenkins 2007; 2008; 2013), I have adhered to a form of metaphysical naturalism that has consequences for epistemology. This is the view that "the world is the way best science says it is (where best science need not be identical to any actual science,

7 See, e.g., Alexander, Mallon, and Weinberg (2010) for discussion of the distinction.

whether present or future), and there's nothing more to be truly said about the world" (Jenkins 2007, 259). Among the *epistemological* naturalisms that I take to be consequences of this more general naturalistic commitment are:

(12) A first-order metaphysical naturalism, according to which there are no epistemically normative facts of a kind irreconcilable with a scientific worldview.

This is reminiscent of Quine's (2), but unlike (2) it is not committal on the question of whether epistemic norms can be reconciled with a scientific worldview. In that respect, it is more like (2*), though unlike (2*) it does not commit to *instrumentalism* about epistemic normativity.

My principal reasons for preferring (12) to (2) are that I do not think we can rule out approaches to epistemic normativity along the lines sketched in Jenkins (2007; 2014a), on which there exist epistemically normative facts *and these are identical to natural facts*. This kind of identification maneuver allows for the naturalistic adoption of a scientific worldview without any pressure to reject epistemic normativity. It also avoids any pressure to recast epistemic normativity as instrumental (as per (2*)), or to metaphysically deflate it (by, for example, adopting an expressivist approach to epistemically normative language).

I'd also be happy to add to (12) something that is also a descendant of a Quinean view:

(1*) A metaphilosophical, normative, methodological naturalism, which says that the methods *relevantly similar to* those of the natural sciences are the only correct methods by which to pursue epistemological inquiry.

Something like (1*) strikes me a natural methodological companion for the view that the world is as best science says it is. The move from Quine's (1) to my (1*), broadening to "methods relevantly similar to those of the natural sciences," is made to allow for the use in epistemology of methods that don't *exactly* resemble those of any other scientific enterprise, but are sufficiently similar, and similar in the right kinds of ways, for them to be recognizably the *sorts* of method that actual or future scientists might employ.

In general, I am inclined to the (approximately Williamsonian[8]) view that inquiry is much more continuous across disciplines and subject matters than the incautious internalization of certain stereotypes of "the sciences," "the humanities," and so on might lead one to think.[9] My sense is that epistemologists do sometimes do squarely *a priori* work, and yet they also do much that is empirically informed, by current science and by all kinds of other academic and extra-academic influences, and they rely (for better and for worse) on the deliverances of what they take to be "common sense." My sense is that all these things are true of scientists too, who, *even according to stereotypes*, draw heavily on the squarely *a priori* discipline of mathematics and use "common sense" *at the very least* as a guide in singling out which hypotheses it is worth their while to subject to empirical testing in the first place. The balance of the various ingredients may vary between different disciplines, but then the balance of the various ingredients also varies (and varies widely) between different subdisciplines of the same discipline.

8 See Williamson (2007). My reasons for thinking interdisciplinary continuity exists are not the same as Williamson's: his include a deep skepticism as to whether the *a priori*/*a posteriori* distinction marks any real difference in kind, which I do not share (for reasons outlined in Jenkins and Kasaki forthcoming).

9 I don't take this to be a form of naturalism *per se*. It is just another view that I hold.

However, I differ significantly from Quine (and, indeed, most epistemological naturalists, many of whom are in some way inspired by Quine) in the area of first-order epistemological naturalisms. It is my contention that committing to epistemological naturalisms like (12) and (1*) need not involve committing to anything like (3), the *rejection of the a priori*. The same goes for more moderate first-order epistemological naturalisms like (7) that aim merely to *restrict* the *a priori*, and for those like (10) that reject or restrict reliance on armchair "intuitions." Nothing in the vicinity of naturalisms (3), (7), or (10) makes my list.

That's partly because none of these first-order epistemological views is, in my opinion, forced on us by the adoption of other kinds of epistemological naturalism like (12) and (1*). Though they are often woven together, first-order epistemological claims like (3), (7), and (10) form a distinct, and separable, strand in the fabric of the naturalistic tradition. It seems to be a popular assumption that there is some obvious connection between adopting a scientific worldview and rejecting or restricting the range of *a priori* and/or armchair knowledge. But I don't think this is the case, unless one is in the grip of the kind of stereotype about what "the sciences" are like that I mentioned before.

In this vein, I have argued (in Jenkins 2013) that naturalism and the *a priori* are far more comfortable bedfellows than is normally assumed. In Jenkins (2013), I consider a range of arguments, all motivated by or premised on forms of naturalism, whose conclusions are meant to be that we should reject or restrict the *a priori*. I consider a Quinean argument in this vein, an argument due to Maddy, and an argument due to Papineau. In each case, it seems to me that the other naturalistic motivations can be preserved without rejecting or restricting the *a priori*. For example, Maddy's naturalistic rejection of a Carnapian two-level conception of inquiry is compatible with believing in *a priori* knowledge, provided that one does not attempt to account for the *a priori* by appealing to Carnapian linguistic frameworks. And this is hardly the only (or even a leading) contender among contemporary accounts of the *a priori* in any case.

In Jenkins (2008), I developed a theory of the *a priori* designed specifically to fit with a kind of naturalism which holds that a broadly scientific worldview is appropriate. On this account, *a priori* knowledge can be secured through reliance on concepts, provided those concepts are epistemically respectable (or *grounded*). When we examine our arithmetical concepts, for example, and thereby arrive at arithmetical beliefs, we are trusting the concepts in question as guides to the nature of arithmetical reality. We need those concepts to be *trustworthy* guides in order for this to be appropriate. The suggestion of Jenkins (2008) is that it is the relationship between concepts and experience that – in cases where things are going well – renders concepts trustworthy. The proposal, in barest outline, is that when our concepts are *sensitive* to experience (and thus to the world as it affects us through experience), those concepts can become trustworthy guides to the world. For more details, readers should refer to Jenkins (2008, ch. 4).

I argue in Jenkins (2013) that, if the account of the *a priori* suggested in Jenkins (2008) (or something close to it) is correct, it follows that all knowledge, whether *a priori* or *a posteriori*, is *ultimately* empirical. And then there is no obvious reason to regard any of it as on a "higher level" than the rest. There is no obvious reason to think, for example, that *a priori* empirical justification cannot be outweighed or undermined by *a posteriori* empirical evidence for a contrary view. There is no obvious reason to doubt that physicists regularly employ *a priori* empirical methodologies or that philosophers regularly employ *a posteriori* ones. There is no Carnapian commitment to the existence of questions that are settled merely by means of our making a (pragmatic) choice of a linguistic framework. All questions,

whether or not *a priori* methods can be used to address them, are to be answered through appropriate sensitivity to empirical input.

There is also a distinctive philosophical tradition that allows for maintaining a robust naturalism alongside a commitment to the *a priori*. This is the Canberra Plan, a cluster of philosophical views associated with philosophers based at the Australian National University during the 1990s and thereafter. The Plan characteristically includes *both* a commitment to *a priori* conceptual analysis as a successful philosophical method *and* a commitment to various kinds of naturalism, including respect for the sciences, the rejection of "unscientific" spooky objects and properties, and materialism.

In fact, so paradigmatic of the Canberra Plan is its combination of naturalism and *a priori*sm that both are covered explicitly in the "The Credo of the Canberra Planners," a lighthearted summary of the views of the Canberra Planners composed by Daniel Nolan (2009), an extract from which reads:

> We believe in conceptual analysis, the a priori, and narrow content.
> Ramsification over platitudes leads to the systematisation of theory,
> and thereby enables identification of the best deservers for theoretical terms.
> We are materialists and functionalists about the mind.
> We reject spooks and epiphenomena of all sorts.
> …
> We respect the opinions of the folk, and are naturalists, with great respect for the findings of science (suitably interpreted).[10]

One further voice on the same side of this issue is that of Rey, who has also argued forcefully against the supposition that naturalistic epistemology commits one to the rejection of the *a priori*. Rey argues that supposed naturalistic arguments against the *a priori* (such as the suggestion that Quinean holism renders it *unnecessary* to posit any *a priori* knowledge and therefore that we *should not* posit any) are unsuccessful, because there are motivations for positing *a priori* knowledge that remain unaddressed by this line of argument. According to Rey, it is in fact an open empirical question whether we have any *a priori* knowledge (see, e.g., Rey 1998).

Since the paradigm Canberra Planners, Rey, and I myself are naturalists (about epistemology and in general) and yet also *a priori*sts, it is not plausible that the rejection of the *a priori* is in any sense a crucial element in every bundle of epistemological views worthy of the name "epistemological naturalism." It is, rather, a separable commitment made by *some* influential epistemological naturalists. (This is not simply a matter of biographical fact; my point here is not just that some of us happen to have this combination of views, but that these combinations of views are currently being developed in viable and interesting ways.)

The same goes for thesis (7), which restricts the applicability of *a priori* methods to the study of natural kinds. On my (naturalistic) account of how the *a priori* works, there is no reason to accept (7). While it strikes me as sensible to doubt that *a priori* methods will be able to tell us *everything there is to know* about natural kinds, I do not see any reason to limit the range of the *a priori* to the nonsubstantive, in this domain or anywhere else.[11] It strikes me that the temptation to restrict the *a priori* in this way may, like

10 A fuller explication of one Canberra Planner's naturalistic *a priori*sm can be found in Jackson (1998).

11 It is less clear to me whether Canberra Planners or Rey can comfortably reject (7).

Maddy's rejection of the *a priori*, simply be harking back to a Carnapian conception of how the *a priori* would have to work.

Lightweight Epistemological Naturalism II: The Armchair

What about thesis (10), the claim that armchair reliance on intuitions is unreliable? In Jenkins (2014b), I begin by addressing the more general issue of what philosophical "intuitions" are, maintaining that philosophers' use of "intuitions" refers to no uniform group of things. However, I argue that on at least one recognizable sense of "intuition" – that on which intuitions are basically *a priori* judgments – various standard objections to reliance on "intuitions" can be resisted. For example, I argue that there is no simple or straightforward argument against reliance on intuitions on the grounds that intuitions are not sufficiently naturalistically respectable. (This, of course, builds on my reasons for thinking it is possible to develop a theory of the *a priori* on which *it* is sufficiently naturalistically respectable.)

"Intuitions" in the Weinberg, Nichols, and Stich (2001) sense are not (or, at least, not only) *a priori* judgments. They are spontaneous judgments about cases. But these too, I would suggest, do not form a uniform kind. Even if we focus attention (as Weinberg, Nichols, and Stich intend) on spontaneous *armchair* judgments about cases of the kind supposedly *characteristic of traditional epistemology*, I would contend that these still do not form a uniform kind. For all Weinberg, Nichols, and Stich say, some of these armchair judgments may be *a priori* judgments that are in perfectly good epistemic standing, while others may be the upshots of unchallenged prejudices based on (inadequate) empirical grounds.

Given its intended reading, then, thesis (10) (which says that "intuitions" are unreliable) is rather like the claim that "judgment" is unreliable, or that "books" are unreliable. Judgment and books *are* unreliable, in the sense that both often get things wrong. And epistemologists have traditionally relied on both. But that's not enough to make for an interesting metaepistemological problem.

A thesis like (1) only becomes interesting or worrying if a case can be made for the further claim that *philosophers are not doing a good enough job* of separating whatever wheat there may be in their spontaneous armchair judgment-making from any surrounding chaff. Making this claim seem plausible is, of course, among the aims of Weinberg, Nichols, and Stich (2001) (they say there is a "substantial body of evidence" which points that way), but it is far from clear that they succeed in doing anything of the sort. Nagel (2012, § 3) provides an excellent breakdown of some of the ways in which the data that Weinberg, Nichols, and Stich collected can be explained without attributing any potentially troubling variation in epistemic intuitions between ethnic groups. She also reports that it was not possible to replicate the Weinberg, Nichols, and Stich results in other trials that were free of the identified confounding factors.

For all that, it may be that there are other data out there which *do* establish that there is something worrying about what epistemologists are or were doing. But what I want to draw attention to here is the fact that whether or not Weinberg, Nichols, and Stich (2001) succeeded in their goals with respect to (7) has nothing particularly to do with how much "naturalism" (in the sense of doctrines like (1*), (12), etc.) it is appropriate to adopt. It has to do with the appropriate interpretation of the specific empirical data that Weinberg, Nichols, and Stich reported. This is a debate that can be (and is being) conducted with self-identified "naturalists" on both sides.

So, doctrines like (10) are again separable from other core naturalistic commitments concerning epistemology. One can be a lightweight epistemological naturalist, in the sense of adopting (1*) and (12), while remaining entirely neutral on the viability, and likely ultimate fate, of the program of negative experimental philosophy.

One upshot of all of this is that maintaining lightweight epistemological naturalism is harmonious with continuing to operate as a philosopher who is often engaged in the kinds of *a priori* and/or armchair inquiry that contribute to the production of work such as the present chapter. According to lightweight epistemological naturalism, there is no tension between the armchair and/or *a priori* nature of (some of) my work and my holding that a broadly scientific worldview is correct and that the application of a broadly scientific methodology is appropriate in philosophy.

Acknowledgments

I am grateful to Chris Meyns, Daniel Nolan, and the editor Kelly James Clark for comments on an earlier draft of this paper, which led to many improvements.

References

Alexander, J., Mallon, R., and Weinberg, J. (2010). Accentuate the Negative. *Review of Philosophy and Psychology* 1: 297–314.

Cappelen, H. (2012). *Philosophy without Intuitions*. Oxford: Oxford University Press.

Colyvan, M. (2009). Naturalising Normativity. In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola, pp. 303–313. Cambridge, MA: MIT Press.

Devitt, M. (2005). There Is No A Priori. In *Contemporary Debates in Epistemology*, edited by Matthias Steup and Ernest Sosa, pp. 105–115. Oxford: Blackwell.

Feldman, R. (2001). Naturalized Epistemology. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/epistemology-naturalized/ (last accessed July 15, 2015).

Goldman, A. (2013). Philosophical Naturalism and Intuitional Methodology. In *The A Priori in Philosophy*, edited by Albert Casullo and Joshua Thurow, pp. 11–44. Oxford: Oxford University Press.

Hare, R.M. (1995). A New Kind of Ethical Naturalism. *Midwest Studies in Philosophy* 20: 340–356.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Jenkins, C. (2007). Epistemic Norms and Natural Facts. *American Philosophical Quarterly* 3: 259–272.

Jenkins, C. (2008). *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*. Oxford: Oxford University Press.

Jenkins, C. (2013). Naturalistic Challenges to the A Priori. In *The A Priori in Philosophy*, edited by Albert Casullo and Joshua Thurow, pp. 274–290. Oxford: Oxford University Press.

Jenkins, C. (2014a). Naturalism and Norms of Inference. In *Naturalizing Epistemic Virtue*, edited by Owen Flanagan and Abrol Fairweather. Cambridge: Cambridge University Press.

Jenkins, C. (2014b). Intuition, "Intuition," Concepts and the A Priori. In *Intuitions*, edited by Anthony Booth and Darrell Rowbottom. Oxford: Oxford University Press.

Jenkins, C. and Kasasi, M. (Forthcoming). The Traditional Conception of the A Priori. *Synthese*.

Kelly, T. (2003). Epistemic Rationality as Instrumental Rationality: A Critique. *Philosophy and Phenomenological Research* 66: 612–640.

Kornblith, H. (1999). In Defense of a Naturalized Epistemology. In *The Blackwell Guide to Epistemology*, edited by John Greco and Ernest Sosa, pp. 158–169. Oxford: Blackwell.

Kornblith, H. (2007). Naturalism and Intuitions. *Grazer Philosophische Studien* 72: 27–49.

Lenman, J. (2006). Moral Naturalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/naturalism-moral/ (last accessed July 15, 2015).

Maddy, P. (2000). Naturalism and the A Priori. In *New Essays on the A Priori*, edited by Paul Boghossian and Christopher Peacocke, pp. 92–116. Oxford: Oxford University Press.

Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research* 85: 495–527.

Nolan, D. (2005). *David Lewis*. Chesham: Acumen.

Nolan, D. (2009). The Credo of the Canberra Planners. Available from: https://sites.google.com/site/professordanielnolan/credo (last accessed July 15, 2015).

Papineau, D. (2007). Naturalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/naturalism (last accessed July 15, 2015).

Paseau, A. (2008). Naturalism in the Philosophy of Mathematics. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/naturalism-mathematics/ (last accessed July 15, 2015).

Plantinga, A. (1996). Methodological Naturalism? In *Facets of Faith and Science*, vol. 1, edited by Jitse van der Meer, pp. 177–221. Lanham: University Press of America.

Quine, W.V.O. (1951). Two Dogmas of Empiricism. *Philosophical Review* 60.Reprinted in Quine, W.V.O. (1980). *From a Logical Point of View: Nine Logico-Philosophical Essays*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W.V.O. (1986). Reply to Morton White. In *The Philosophy of W.V.O. Quine*, edited by L. Hahn and P. Schlipp, pp. 663–665. La Salle: Open Court.

Rey, G. (1998). A Naturalistic A Priori. *Philosophical Studies* 92: 25–43.

Stoljar, D. (2009). Physicalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/physicalism/ (last accessed July 15, 2015).

Weinberg, J., Nichols, S., and Stich, S. (2001). Normativity and Epistemic Intuitions. *Philosophical Topics* 29: 429–460.

Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Oxford University Press.

17

Dewey, Naturalism, and the Problem of Knowledge

DOUGLAS MCDERMID

> The question whether objective truth can be attributed to human thinking is not a question of theory but is a practical question. Man must prove the truth – i.e. the reality and power, the this-sidedness of his thinking in practice. The dispute over the reality or non-reality of thinking that is isolated from practice is a purely *scholastic* question.
>
> Karl Marx (1845)

I

Is epistemology the opium of modern philosophers, numbing fine minds and filling them with world-denying dreams? The American pragmatist John Dewey (1859–1952) was convinced that it was, and he thought his brand of naturalism was the proper way to treat this pernicious addiction. In this chapter, I shall present Dewey's audacious and influential defense of this thesis. Audacious, because Dewey attacks assumptions about the nature of mind, knowledge, truth, and reality which have long been central to Western philosophy. Influential, because Dewey's bold polemic has served as a major reference point for several leading pragmatists – Sidney Hook (1902–1989), W.V.O. Quine (1908–2000), Richard Rorty (1931–2007), and Philip Kitcher (b. 1947) – whose naturalist predilections are well-known.

II

According to Dewey, Plato was dead wrong: there is no way to escape from the tenebrous cave of one's culture, no way to behold a radiant Reality of which *hoi polloi* know nothing. Yet if philosophy has no privileged and ahistorical knowledge of the True and the Good and the Beautiful, Dewey reasons, its practitioners must drop their Platonic pretensions and admit that they do not possess any absolute or final standards by which the institutions and practices of their culture can be judged. Instead of continuing to pose as bloodless spectators

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

of all time and existence, philosophers must henceforth acknowledge that there is nothing – no realm of perfect Being, no supersensible world of Forms, no noumenal domain apprehended via intellectual intuition, no body of synthetic *a priori* truths about Truth or Reality – which it is their prerogative to grasp or limn; and they must also acknowledge that there is no Archimedean point or God's-eye perspective from which they can pass judgment on the methods or achievements of scientific inquiry. Philosophy is constrained by science as well as continuous with it, Dewey insists, and philosophers who treat scientists with lordly condescension only make themselves look foolish:

> [Philosophy] has no call to create a world of "reality" *de novo*, nor to delve into secrets of Being hidden from common-sense and science. It has no stock of information or body of knowledge peculiarly its own; if it does not always become ridiculous when it sets up as a rival of science, it is only because a particular philosopher happens to be also, as a human being, a prophetic man of science. Its business is to accept and to utilize for a purpose the best available knowledge of its own time and place. And this purpose is criticism of beliefs, institutions, customs, policies with respect to their bearing upon good. This does not mean their bearing upon *the* good, as something itself attained and formulated in philosophy. For as philosophy has no private score of knowledge or of methods for attaining truth, so it has no private access to good. As it accepts knowledge of facts and principles from those competent in inquiry and discovery, so it accepts the goods that are diffused in human experience. It has no Mosaic nor Pauline authority of revelation entrusted to it. But it has the authority of intelligence, of criticism of these common and natural goods. (Dewey 1929, 330; cf. 1917/1981, 89; 1944, 1–2, 9)

Once we accept that philosophy is constrained by science, however, we must make room for the possibility that philosophy's "timeless" or "perennial" questions are actually nothing of the kind. For we may well discover that at least some of these problems are mere cultural survivals or hangovers from the past; problems, that is to say, whose underlying assumptions may have satisfied the best minds of the age in which they were originally formulated, but which cannot be reconciled with what we are now entitled to think we know about the world and our place in it. And the problems which define "the whole modern industry of epistemology" (Dewey 1910, 297; cf. 1917/1981, 75) are unfit survivals of precisely this sort, according to Dewey. Instead of treating them with respect on account of their longevity, he thinks, we need to get rid of them on account of their artificiality; instead of trying to answer them, we need to stop asking them. The Gordian knot of epistemology, in short, must not be unraveled; it must simply be cut, swiftly and cleanly.

Do we have the right to turn our backs on tradition in this way? Of course we do, replies Dewey; since longevity is no guarantee of relevance, the mere fact that a problem is traditional does not mean that we have an obligation to solve it. Besides, philosophical breakthroughs typically occur only when philosophers have the courage to let the old unsolved enigmas fade away:

> Old ideas give way slowly; for they are more than abstract logical forms and categories. They are habits, predispositions, deeply engrained attitudes of aversion and preference. Moreover, the conviction persists – though history shows it to be a hallucination – that all the questions that the human mind has asked are questions that can be answered in terms of the alternatives that the questions themselves present. But in fact intellectual progress usually occurs through sheer abandonment of questions together with both of the alternatives they assume – an abandonment that results from their decreasing vitality and a change of urgent interest. We do

> not solve them; we get over them. Old questions are solved by disappearing, evaporating, while new questions corresponding to the changed attitude of endeavor and preference take their place. (Dewey 1910, 19; cf. 1917/1981, 58–59, 76–77; 1930/1981, 13)

What makes Dewey so sure that epistemology rests on assumptions which we can no longer accept, and what does he think will be gained for philosophy if epistemology is abandoned? The answers to these questions are contained in his main argument against epistemology-centered philosophy. According to my reconstruction, this argument has four premises.

Premise A

Modern philosophy is epistemology-centered, and its agenda has been profoundly shaped by the idea that the theory of knowledge is first philosophy. Understanding what knowledge is, analyzing knowledge's structure, stating the conditions of its possibility, identifying its sources, determining the constitution of its objects, defining its limits: these issues have dominated philosophical inquiry since the 17th century, and there is widespread agreement about their foundational status. Indeed, it is now almost a truism to say that philosophical questions about the nature of knowing are logically prior to questions about the nature of being, and that we have no right to address the latter until we have answered the former.

> Everybody knows that the trend of modern philosophy has been to arrive at theories regarding the nature of the universe by means of theories regarding the nature of knowledge – a procedure which reverses the apparently more judicious method of the ancients in basing their conclusions about knowledge on the nature of the universe in which knowledge occurs. (Dewey 1929/1990, 33)

> It is a commonplace that the chief divisions of modern philosophy, idealism in its different kinds, realisms of different brands, so-called common-sense dualism, agnosticism, relativism, phenomenalism, have grown up around the epistemological problem of the general relation of subject and object. Problems not openly epistemological, such as whether the relation of changes in consciousness to physical changes is one of interaction, parallelism, or automatism have the same origin. (Dewey 1917/1981, 76)

The defining problem of epistemology, its logical center of gravity and thematic core, is the problem of human knowledge in general, of "knowledge *überhaupt*" (Dewey 1917/1981, 76; cf. 1910, 97). Roughly speaking, this is the problem of understanding how we can have any knowledge of an objective world existing independently of the mind if it is assumed that we have direct epistemic access to nothing but the purely subjective contents of our own minds: ideas, impressions, sense data, sensations, representations, experiences, or perceptions. It is worth remembering that this problem, far from being a perennial riddle, acquired its canonical status through the writings of Descartes in the mid-17th century. Nothing quite like this Cartesian query was on the agenda of the ancient Greeks or the medieval Schoolmen, most of whom would almost certainly have found it perverse, outlandish, or pointless. For the record, this is more or less how Dewey himself sees the problem of the external world; yet his reasons for taking such a dim view of it, it must be stressed, are also reasons for taking a dim view of the way in which knowledge itself has been understood in Western philosophy since the Greeks. But more of this later, when we come to Premise D.

Premise B

The formulation of the Cartesian problem of knowledge presupposes a host of dualisms and dichotomies: between mind and world, subject and object, appearance and reality, experience and nature, internal and external, private and public. In effect, the task of epistemologists is to find an argument that will build a sturdy knowledge-bearing bridge leading from the first term to the second term of each dualism – viz., from mind to world, from the subjective to the objective, from appearance to reality, and so on. If we can invent or discover such an argument, well and good; we will have shown that human beings are not victims of their own subjectivity, forever imprisoned inside their own minds. If we cannot come up with such an argument, however, we must concede that each of us has been sentenced for life to the epistemological equivalent of solitary confinement, and that human knowledge is nothing but an inmate's perfervid dream of freedom.

> The discipline termed Epistemology assumes, rightly or wrongly, a self-enclosed island of mind on one side, individual and private and only private; over against this is set a world of objects which are physically or cosmically there – and only there. Then it is naturally worried about how the mind can get outside of itself to know a world beyond, or how the world out there can creep into "consciousness." (Dewey 1911/1978, 18)
>
> The problem of knowledge *überhaupt* exists because it is assumed that there is a knower in general, who is outside of the world to be known, and who is defined in terms antithetical to the traits of the world. (Dewey 1917/1981, 76)

Premise C

The problem of the external world is not merely difficult; it is, in truth, insoluble. Think of it this way: if we have immediate knowledge of nothing but our mental representations, we can never compare our representations with the extra-mental reality they purport to represent; and if we can never compare our thoughts directly with things, we can never know whether our worldview corresponds or conforms to the world: "How can anybody look at *both* an object (event) and a proposition about it so as to determine whether the two 'correspond'?" (Dewey 1946, 352, cf. 343). If this argument succeeds, it must be concluded that there is no way for us to build a bridge between mind and world, to close the gap between subject and object, to connect appearance and reality. Once these contrasting terms have been prized apart and isolated from each other, philosophy cannot reestablish continuity between them, and it cannot prove that the spheres in question must correspond on any level.

> When objects are isolated from the experience through which they are reached and in which they function, experience itself becomes reduced to the mere process of experiencing, and experiencing is therefore treated as if it were also complete in itself. We get the absurdity of an experiencing which experiences only itself, states and processes of consciousness, instead of the things of nature. Since the seventeenth century this conception of experience as the equivalent of subjective private consciousness set over against nature, which consists wholly of physical objects, has wrought havoc in philosophy. It is responsible for the feeling…that "nature" and "experience" are names for things which have nothing to do with each other. (Dewey 1929, 13)
>
> As long as theories of knowledge are framed in terms of organs assigned to mind or consciousness, whether sense or reason or any combination of the two, organs occupied, it is alleged, in reproducing or grasping antecedent reality, there will continue to exist such generalized

> skeptical philosophies…Theories which assume that the knowing subject, that mind or consciousness, have an inherent capacity to disclose reality, a capacity operating apart from any overt interactions of the organism with surrounding conditions, are invitations to general philosophical doubt. (Dewey 1929/1990, 154)

Hence the very moment we endorse the dualisms and dichotomies which are built into epistemology, we are faced with an insoluble problem. This means that skepticism is, at the very least, conditionally or hypothetically correct: *if* the problem of the external world is a real problem, *if* its basic presuppositions are indeed beyond reproach, then an honest and competent philosopher has no choice but to admit that human knowledge is an ubiquitous illusion, and that Nature is a Something-We-Know-Not-What. In this sense, the task that epistemologists have set for themselves is as hopeless and dreadful as the infernal punishments found in Greek mythology: "The ancient myth of Tantalus and his effort to drink the water before him seems to be ingeniously prophetic of modern epistemology" (Dewey 1910, 186, cf. 257).

Premise D

There is a way out, however, because the representationist dualisms presupposed by the problem of knowledge are untenable.

> Familiarity breeds credulity oftener than contempt. How can a problem be artificial when men have been busy discussing it almost for 300 years? But if the assumption that experience is something set over against the world is contrary to fact, then the problem of how self or mind or subjective experience or consciousness can reach knowledge of an external world is assuredly a meaningless problem. Whatever questions there will be about knowledge, they will not be the kinds of problem which have formed epistemology.
>
> The problem of knowledge as conceived in the industry of epistemology is the problem of knowledge in general – of the possibility, extent, and validity of knowledge in general. What does this "in general" mean? In ordinary life there are problems a-plenty of knowledge in particular; every conclusion we try to reach, theoretical or practical, affords such a problem. But there is no problem of knowledge in general. (Dewey 1917/1981, 75; cf. 1916, 279)

The epistemologist's question, "How is knowledge of the external world possible?", does not deserve an answer because it is not a real question. Why is it not a real question? Because it rests on assumptions about the nature of mind, knowledge, truth, and reality which no contemporary naturalist can be expected to endorse, inasmuch as those assumptions "have no empirical standing" (Dewey 1917/1981, 77). The label that Dewey slaps on this constellation of old-fashioned assumptions is *the spectator theory of knowledge*, and in Section III we shall examine what he has to say about it.

Conclusion

Philosophy can no longer be centered on epistemology; for once we have seen through the spectator theory of knowledge, we have seen through the problem of knowledge *überhaupt*. The question about knowledge which philosophers should ask is not whether our ideas stand in a mysterious and static relation of "correspondence" with things-in-themselves to which we have no representation-independent access. Not only is that question unanswerable,

it is also not even worth asking; for it is "a purely scholastic question" (Marx 1845), an idle riddle without any practical implications. No, the questions about knowledge which philosophers should ask – and these are questions of practice, not of unearthly theory – are these: Do our ideas actually work? Do they give us dominion over Nature? Can we use them to remake reality in our own image? To ask these questions instead of the skeptic's question is to recommend that philosophy become less Cartesian and more Baconian; less concerned with knowledge as representational fit, that is, and more concerned with knowledge as the power to better man's estate. And this recommendation will be welcomed by all who maintain, as Dewey does, that philosophy's function is not to survey The World objectively, but to improve our world practically – to deal with the "problems of men" (Dewey 1917/1981, 95; 1946) as best we can, employing the self-correcting method of intelligence that has worked so well in science.

III

We now need to work through Dewey's answers to three questions raised by Premise D. First, what exactly is meant by "the spectator theory of knowledge"? Second, what are the basic assumptions behind the spectator theory? Third, why should that theory be rejected?

1

The spectator theory models our understanding of knowledge on our prescientific understanding of vision. According to this model, knowing is analogous to the act of seeing physical objects from a detached and disengaged point of view. To grasp what this entails, consider a simple example. When I look at yonder tree, is the tree itself changed in any way by my gaze? Assuredly not; but *I* have been changed, inasmuch as my condition is not now what it was before I looked at the tree and had certain experiences. In the same way, we are told, the knower is transformed by the acquisition of knowledge, whereas the objects of knowledge are such that they cannot be altered by being known: "The real object is the object so fixed in its regal aloofness that it is a king to any beholding mind that may gaze upon it" (Dewey 1929/1990, 19). At bottom, then, knowledge is essentially a matter of representation or duplication: objects of knowledge are passively apprehended or beheld by a knowing subject who does not interact with them overtly or add anything to them, but merely mirrors, reflects, or reproduces what is already there.

In one form or another, this powerful intuition has informed practically all significant thinking about knowledge in Western philosophy. Whether we are talking about Plato (whose immutable Forms are the objects of knowledge and contemplation par excellence), Aquinas (for whom the highest form of knowledge was the vision of God, who is pure actuality), Locke (whose passively received simple ideas are the building blocks of knowledge), Kant (who was forced to deny we had knowledge of things in themselves because human knowledge imposes conditions on its objects which are foreign to their nature), or logical positivists like Moritz Schlick (who held that our knowledge of things ultimately rests on a bare, unconceptualized Given), sooner or later we stumble upon some version of the idea that true knowledge requires pure receptivity and maximal detachment on the part of the knower, and a fixed and unchanging constitution on the part of the object known.

> Special theories of knowledge differ enormously from one another. Their quarrels with one another fill the air. The din thus created makes us deaf to the way in which they say one thing in common…They all hold that the operation of inquiry excludes any element of practical activity that enters into the construction of the object known…The common essence of all these theories, in short, is that what is known is antecedent to the mental act of observation and inquiry, and is totally unaffected by these acts; otherwise, it would not be fixed and unchangeable. This negative condition, that the processes of search, investigation, reflection, involved in knowledge relate to something having prior being, fixes once for all the main characters attributed to mind, and to the organs of knowing. They *must* be outside what is known, so as not to interact with the object to be known. If the world "interaction" be used, it cannot denote that overt production of change it signifies in its ordinary and practical use. (Dewey 1929/1990, 18–19)

> [T]he counterpart of the invidiously real reality is the spectator notion of knowledge. If the knower, however defined, is set over against the world to be known, knowing consists in possessing a transcript, more or less accurate but otiose, of real things. Whether this transcript is presentative in character (as realists say) or whether it is by means of states of consciousness which represent things (as subjectivists say), is a matter of great importance in its own context. But, in another regard, this difference is negligible in comparison with the point in which both agree. Knowing is viewing from outside. (Dewey 1917/1981, 91, cf. 75)

2

The spectator theory of knowledge has its roots in the separation of theory from practice, and this separation in turn has its roots in what Dewey dubbed *the quest for certainty*. This phrase may immediately make us think of Descartes and his quixotic pursuit of indubitable principles, but Dewey is after much bigger game than a single thinker, even if the thinker in question has been hailed as the father of modern philosophy. For the quest for certainty, Dewey assures his readers, existed long before there was such a thing as philosophy.

Early human beings, we are told, sought certainty and prized it. Why? Was it because they desired knowledge for its own sake, or because they pursued truth as an end in itself? Not at all; what prompted the quest for certainty, Dewey claims, was a pressing practical problem which is at least as old as *Homo sapiens*: that of trying to find a measure of safety and security in a world which is fraught with peril and thick with risk: "Man who lives in a world of hazards is compelled to seek for security" (Dewey 1929/1990, 3; cf. 1922, 236–238). Our prehistoric ancestors faced this problem in its most acute and dramatic form, he adds, since they were ignorant of the methods which yield knowledge of the order of things. Because their understanding of this order was remarkably crude and rudimentary, they found Nature to be anything but natural:

> [P]rimitive men had none of the elaborate arts of protection and use which we now enjoy and no confidence in his own powers when they were reinforced by appliances of art. He lived under conditions in which he was extraordinarily exposed to peril, and at the same time he was without the means of defense which are to-day matters of course. Most of our simplest tools and utensils did not exist; there was no accurate foresight; men faced the forces of nature in a state of nakedness which was more than physical; save under unusually benign conditions he was beset by dangers that knew no remission. In consequence, mystery attended experiences of good and evil; they could not be traced to their natural causes and they seemed to be the dispensations, the gifts and the inflictions, of powers beyond possibility of control. The precarious crises of birth, puberty, illness, death, war, famine, plague, the uncertainties of the hunt, the vicissitudes of climate and the great season changes, kept imagination occupied with the uncertain. (Dewey 1929/1990, 8; cf. 1934, 24–25, 56)

Unable to control natural phenomena or to anticipate the course of events, Dewey's "primitive men" were at Nature's mercy. Since Nature is not especially merciful, it is hardly surprising that men and women, lacking the power to improve the unforgiving conditions under which they lived, sought consolation in a more congenial inner world of imagination and myth. And this retreat from reality, Dewey declares, eventually led to the development of religion. Instead of trying to change a world which they could not fathom, human beings changed themselves; instead of seeking to triumph over Nature, they cultivated an attitude of patient resignation, and submitted to the will of gods whom they credited with the very powers they desired for themselves. The essential point here is that the effective arts of practical control which we take for granted – agriculture, hunting, cooking, weaving, building, metal-working, weapon-making, navigation, medicine, engineering, and so on – were not, so to speak, given to Adam by God in the Garden of Eden; instead, they were discovered through a long and painful process of trial and error. Before these arts were adequately developed, most human beings had little or no faith in the power of practical activity to improve their lives in any meaningful way, and we can hardly blame them. Surrounded by signs of change and decay, weary of failure and frustration, irked by the aleatory and protean character of their existence, they found solace in the realm of religion because it offered substitute satisfactions. When philosophy finally arrived on the scene, these substitutes assumed a less overtly mythological form; yet the underlying motivations were the same: the desire to escape change and chance, to find a source of otherworldly peace that could not be lost by accident or through action. Thus, the quest for certainty was something which philosophy inherited from religion; it was not something which philosophy invented on its own.

Now, how did the philosophical version of the quest for certainty lead to the spectator theory of knowledge? Since it was recognized that practice could never deliver absolute certainty, Dewey explains, it was agreed that certainty could be found only in the sphere of knowledge – but knowledge of what? There's the rub; for if we confine ourselves to the familiar realm of flux and impermanence, to the world of endless change and ceaseless becoming, it is evident that no final assurances or indubitable truths can be had. The objects of true knowledge must therefore be located in another realm or domain, one not infected with the inconstancy and unpredictability that characterizes the ephemeral phenomena with which our practice deals. In other words, knowledge's proper objects must be untainted by change and replete with pure Being; they must be immutable and fixed, constant and complete in themselves, actual and static, final and fully real. And because they are immutable, the objects of knowledge cannot be changed by being known any more than a tree is changed by being seen. True knowledge can therefore make no difference to what is known; for the knower merely beholds objects of knowledge from afar like a cosmic voyeur, or submits to them religiously in the hope of being delivered from the wild and incomprehensible contingencies of life. To know, it was concluded, is to let reality affect us without affecting reality; it is to contemplate a ready-made world from the perspective of a detached and inert spectator whose satisfaction is essentially private and aesthetic.

In the 17th century, Dewey claims, this way of thinking about knowledge and its objects gave rise to the modern epistemological dualisms between mind and world, subject and object, appearance and reality, experience and nature. For it was around that time that philosophers in Europe, rightly impressed with both the progress and the promise of the new science, began to ask: What is the nature of our knowledge of Nature? Assuming the

spectator theory of knowledge is correct, Nature can be known only if the knower is altogether outside of its domain; otherwise, the knower might alter things in the very act of knowing them, which would mean that knowledge would be self-defeating. In this way, philosophers reached the conclusion that Nature can be known only if the seat and organs of knowledge lie in a non-natural sphere; that is, in a domain that is isolated from the world of ordinary practical activity and defined in sharp opposition to it. And this is more or less how the vast majority of modern philosophers in the West have thought of the mind down to the present day, according to Dewey: as a realm of pure subjectivity that knows an "external world" by representing or reproducing elements of it in the private medium of consciousness.

3

The rigid separation between theory and practice that produced the spectator theory is a direct expression of cultural conditions which no longer obtain. Theory was originally elevated over practice, it has already been suggested, partly because our ancestors identified the limitations of their practices with the limitations of practice per se, and because they found in the serene and self-contained joys of pure thought abundant compensation for their failure to subdue Nature. In addition, they were influenced by religious interpretations of life in which Nature was portrayed as fallen and corrupt, as a dreary vale of tears and a deadly snare for unwary souls, a realm where all was vanity and nothing under the sun was new. Since no solid or lasting satisfaction can be found on earth, it was said, the present life must be seen as a state of probation, a wilderness through which humans must pass on their grim pilgrimage to glory everlasting. Inspired by this pessimistic view of the spatiotemporal world, sensitive men and women turned away from the things of ordinary experience in search of a higher and better reality, where their restless hearts could find enduring peace and their souls endless felicity.

Class differences, too, undoubtedly played a significant role in promoting the separation of theory and practice. The refined gentleman of classical antiquity, it is plain, seldom got his soft hands dirty; it was his social and economic inferiors, the unlettered tillers of the soil and the calloused manipulators of matter, who were forced to live by the sweat of their brow and to earn their bread by doing his bidding. In other words, contemplation was reserved for the leisured and powerful elite, whereas manual labor was the exclusive province of artisans, servants, slaves (Aristotle's "living tools"), and unlettered drudges. In pre-liberal societies, therefore, philosophy's self-serving elevation of theory over practice not only reflected existing power relations, but also reinforced them, and helped preserve the status quo by providing an ideological justification for various forms of inequality and exploitation.

Why does Dewey call our attention to these things? His point is a skeptical one: the genesis and propagation of the ideas that spawned the spectator theory of knowledge can be accounted for in terms of a host of extraphilosophical factors – religious, social, political, economic, and so on – which have nothing to do with whether those ideas are philosophically defensible. This does not show that these ideas are false, of course; but it should lead us to ask whether they are well-founded, since their remarkable persistence and popularity can no longer be taken as clear and unambiguous evidence of their truth. Instead of regarding them as sacrosanct or self-evident givens, we now have the right – nay, the duty – to ask whether philosophers should continue to endorse them and use them.

It is at this point that Dewey, putting his cards on the table, tells his readers that recent cultural developments have discredited the spectator theory and its presuppositions. What sorts of development does he have in mind? Here are six examples:

(1) Though the idea that theory is inherently superior to practice might have seemed plausible when the arts of prediction and control were largely undeveloped, the last three centuries have shown us that practical activity can greatly improve our lives, and that science and its myriad applications (e.g., medicine, engineering, industry, technology, and so on) can definitely lead to the betterment of man's estate. Why, then, should we still bow before the old burnished idol of theoretical certainty? Such certainty, it will be recalled, was originally a substitute which our desperately anxious ancestors sought because they lacked what we possess: power over Nature.

(2) The history of natural science, which reminds us how the certainties of one generation become the myths of a later age, clearly supports a robust form of fallibilism: the doctrine that no theory, judgment, or observational statement can ever be reckoned absolutely final or beyond revision. Since philosophy, too, has given us good reason to embrace this sort of fallibilism, the quest for certainty must be seen for the intellectual dead end that it is. It is time we admitted to ourselves that our worldview, like the world itself, is always a work in progress; that there are no adamantine guarantees, no irrefragable truths; that hypotheses are all we will ever have; and that hypotheses become stumbling blocks when they are converted into dogmas.

(3) The methods and procedures of modern science have completely revolutionized our philosophical understanding of the relation between theory and practice, knowledge and conduct, thinking and doing. Since knowledge in science is the product of well-conducted inquiry, and since inquiry involves overt action in the form of experimentation, naturalists must conclude that knowers are agents who make a concrete difference in the world, not disembodied demigods who contemplate it with Olympian detachment. Because we acquire knowledge of things by transforming them practically, knowing must be seen as a natural and public process, and not as an episode occurring in a mysterious inner kingdom called "the mind."

(4) It should also be noted that observation in science is by no means spectatorial or subjective, disinterested or detached. Instead of receiving data passively and indiscriminately like a Lockean tabula rasa, observers are active and ruthlessly selective. Like treasure-hunters or explorers, they are typically *looking for something* – the answer to a question, the solution to a problem, evidence for or against a particular theory or hypothesis – and their search is guided by a background set of interests and expectations. Since to observe is to extract facts (with instruments, if need be), observation is always intervention – a point underscored by Heisenberg's uncertainty principle in physics.

(5) According to the classical tradition in Western philosophy, knowledge's true and proper objects are the inner natures of mutable things; that is, their forms or essences. Such essences, it was held, were fixed and final; and their fixity and permanence conferred upon them the intrinsic intelligibility which objects of knowledge were expected to possess. Unfortunately for the classical tradition, Darwin showed that the inner natures of things are the children of time and chance, and that Nature itself is a work in progress, without any preordained telos or endpoint. The moral we should draw, if

we are naturalists, is obvious: the objects of knowledge can no longer be identified with what is absolutely static or unchanging.

(6) If we are good naturalists, we shall be good Darwinians; and if we are good Darwinians, we shall think of ourselves as organisms striving constantly to adapt and adjust to a complex and ever-changing environment. This biological turn brings about a radical change of perspective in philosophy, for it forces us to revise our inherited understandings of experience, knowledge, and truth. Since the experience of an organism is an active and practical transaction, human experience can no longer be regarded as a passive or subjective affair; and since knowledge understood in natural and biological terms is instrumental and not an end in itself, human knowledge can no longer be regarded as a perfect mirror held up to Nature. Ideas must now be seen as plans of action; true ideas, as plans of action that work best in the long run. True ideas must correspond to reality, therefore, only in a functional or operational sense: the sense in which a solution corresponds to a specific problem, or in which a tool corresponds to a particular task.

IV

The bottom line? Once we accept naturalism, we can no longer engage in epistemology with a clean intellectual conscience; for the spectator theory of knowledge and the separation of theory and practice will not recommend themselves to any philosopher who is sensible and self-effacing enough to deny that philosophy is prior to science. In other words, Dewey's main negative contention – and this leads us back, as it should, to Premise D – is that "the epistemological universe of discourse" (Dewey 1917/1981, 92) must be regarded as mythological by philosophers who have broken with the spirit (and not merely the letter) of supernaturalism.

Under the sobering and salubrious influence of naturalism, Dewey predicts, philosophy will be different: less self-absorbed and more benevolent, less dogmatic and more experimental, less wedded to the status quo and more interested in the future, less inclined to endorse thought-blocking dualisms and more inclined to recognize continuities. Freed from otherworldly aspirations and an attachment to barren inherited conundrums, philosophers can finally devote themselves wholly to the practical task of creating a better, more democratic world: a world pregnant with hitherto unrealized possibilities of meaning and joy and true community, where "making a living economically speaking will be at one with making a life that is worth living" and "poetry and religious feeling will be the unforced flowers of life" (Dewey 1920, 164; cf. 1922, 330–332). This, at any rate, was Dewey's buoyant and bracing Whitmanesque hope, and his commitment to philosophical naturalism cannot be understood apart from it.

Further Reading

For further information on the topics raised in this chapter, see Santayana (1925), Hook (1939; 1961; 2002), Rorty (1979; 1982), West (1989), Ryan (1995), Kloppenberg (1996), Menand (2001), Godfrey-Smith (2002), Kim (2003), McDermid (2006), Kitcher (2012), and Lachs (2012).

References

Dewey, J. (1910). *The Influence of Darwin on Philosophy and Other Essays*. Amherst: Prometheus Books.

Dewey, J. (1911/1978). The Problem of Truth. In *John Dewey: The Middle Works (1899–1924)*, vol. 6, edited by J.A. Boydston. Carbondale and Edwardsville: Southern Illinois University Press.

Dewey, J. (1916). *Essays in Experimental Logic*. Chicago: University of Chicago Press.

Dewey, J. (1917/1981). The Need for a Recovery of Philosophy. In *The Philosophy of John Dewey*, edited by J.J. McDermott. Chicago: University of Chicago Press.

Dewey, J. (1920). *Reconstruction in Philosophy*. New York, Mentor Books.

Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York: The Modern Library.

Dewey, J. (1929). *Experience and Nature*. LaSalle: Open Court.

Dewey, J. (1929/1990). *The Quest for Certainty: A Study of the Relation between Knowledge and Action*. In *John Dewey: The Late Works (1925–1953)*, vol. 4, edited by J.A. Boydston. Carbondale and Edwardsville: Southern Illinois University Press.

Dewey, J. (1930/1981). From Absolutism to Experimentalism. In *The Philosophy of John Dewey*, edited by J.J. McDermott. Chicago: University of Chicago Press.

Dewey, J. (1934). *A Common Faith*. New Haven: Yale University Press.

Dewey, J. (1944). Anti-Naturalism in Extremis. In *Naturalism and the Human Spirit*, edited by Y.H. Krikorian. New York: Columbia University Press.

Dewey, J. (1946). *Problems of Men*. New York: Philosophical Library.

Godfrey-Smith, P. (2002). Dewey on Naturalism, Realism and Science. *Philosophy of Science* 69: 1–13.

Hook, S. (1939). *John Dewey: An Intellectual Portrait*. Amherst: Prometheus Books.

Hook, S. (1961). *The Quest for Being, and Other Studies in Naturalism and Humanism*. London: Macmillan.

Hook, S. (2002). *Sidney Hook on Pragmatism, Democracy, and Freedom: The Essential Essays*, edited by Robert B. Talisse and Robert Tempio. Amherst: Prometheus Books.

Kim, J. (2003). The American Origins of Philosophical Naturalism. *Journal of Philosophical Research* 28: 83–98.

Kitcher, P. (2012). *Preludes to Pragmatism: Toward a Reconstruction of Philosophy*. Oxford: Oxford University Press.

Kloppenberg, J.T. (1996). Pragmatism: An Old Name for some New Ways of Thinking? *Journal of American History* 83(1): 100–138.

Lachs, J. (2012). *Stoic Pragmatism*. Bloomington: Indiana University Press.

Marx, K. (1845). *Theses on Feuerbach*. Available from: http://www.marxists.org/archive/marx/works/1845/theses/theses.htm (last accessed July 15, 2015).

McDermid, D. (2006). *The Varieties of Pragmatism: Truth, Realism, and Knowledge from James to Rorty*. New York: Continuum.

Menand, L. (2001). *The Metaphysical Club*. New York: Farrar, Straus and Giroux.

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rorty, R. (1982). *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.

Ryan, A. (1995). *John Dewey and the High Tide of American Liberalism*. New York: Norton.

Santayana, G. (1925). Dewey's Naturalistic Metaphysics. *Journal of Philosophy* 22(25): 673–688.

West, C. (1989). *The American Evasion of Philosophy: A Genealogy of Pragmatism*. Madison: University of Wisconsin.

18

Truth and Naturalism

DOUGLAS EDWARDS, FILIPPO FERRARI, AND
MICHAEL P. LYNCH

Introduction

In this chapter we address the issue of whether truth is a natural phenomenon. We take this to be an important question for both those working on truth and those working on naturalism. For theorists of truth, answering the question of whether truth is natural will tell us more about the nature of truth (or lack of it), and the relations between truth and other properties of interest. To those who wish to have truth as part of the natural order, answering this question is of paramount importance. We will focus primarily on the kinds of theory of truth that occupy the central positions in current debates about truth, namely correspondence theories, deflationary theories, epistemic theories, and pluralist theories, and aim to discern the extent to which truth is a natural property on each view. We will close by making some brief remarks on how the possible normativity of truth bears on the question of whether truth is a natural property.

Two Notions of "Natural Property"

Before talking about truth, we should say a bit about how we understand the term "natural property." We will not go into the complex issues in this area in great depth, as we trust that they will be extensively discussed elsewhere in this volume, but a little discussion is necessary to clarify the ways in which we will evaluate theories of truth in relation to naturalism. There are two distinct notions of "natural property" that we wish to consider and use going forward. The first is what we call the "traditional" conception of naturalness ("T-natural"), and the second is what we will call the "Lewisian" conception of naturalness ("L-natural"). As both of these understandings of "natural property" have significant influence in contemporary debates in metaphysics, we feel it is worthwhile to assess theories of truth in terms of both.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

T-Natural

The first notion of naturalness we will consider is what we can call T-naturalism. While a slippery notion, it can be put in its simplest terms as the view that science determines what exists. In other words, a natural entity, or a natural property, will be something that is accountable in scientific terms. This is a vague definition, particularly until we say a bit more about what science is taken to be. The most prominent naturalist view of this sort in recent times takes it that everything that exists is – in some sense – physical, and that physics is the most basic of all scientific disciplines. On this view, a natural entity, or a natural property, will be a property that either is a physical property or depends on physical properties. Such properties are placed within the empirical realm of causal interaction, and naturalism of this sort states that a property is natural if it is a physical property, open to empirical investigation, and plays some role in the causal push-and-pull of the physical world. As there are some issues about how exactly to understand "physical," we will take a somewhat moderate stance on this physicalist aspect of naturalism, and focus primarily on the following two features, with the assumption that exhibiting them brings with it an entailment to be either a physical property or a property that strongly supervenes on the physical:

(1) Natural properties are open to empirical investigation.
(2) Natural properties have causal powers.

We take it that this is a fairly uncontroversial, if perhaps simplified, statement of T-naturalism.

Of course, a further feature, sometimes thought to be essential to T-naturalism, is the idea that natural properties are non-normative. In what follows we will remain neutral about whether non-normativity is a key feature of T-natural properties, leaving our discussion of the question of the normativity of truth to the final section.

L-Natural

The second notion of naturalness that we will consider comes from David Lewis; it has garnered considerable interest in contemporary work in metaphysics. We will use "L-natural" to refer to this notion, and we take the following quote as our starting point:

> Sharing of [the natural properties][1] makes for qualitative similarity, they carve at the joints, they are intrinsic, they are highly specific, the sets of their instances are *ipso facto* not entirely miscellaneous, there are just enough of them to characterise things completely and without redundancy. (Lewis 1986, 228)

There are two[2] central ideas mentioned in Lewis's quote that we wish to focus on:

(1) Natural properties mark similarities between bearers.
(2) Natural properties are fundamental.

1 Lewis is actually talking about sparse properties here, but, as he takes natural properties to play the role of sparse properties, I think we are licensed to assume that natural properties also exhibit these features. This also fits with his talk of the features of natural properties throughout Lewis (1983).

2 Lewis also mentions the notion of intrinsicality, but, given that we are interested in *truth* as a natural property, we can disregard this feature, as it is implausible to think that beliefs are true or false intrinsically. We do not think that this is a serious problem, as these features are plausibly thought of as marks of L-naturalness, as opposed to necessary and sufficient conditions, and, as noted in the main text, the notion comes in degrees.

There is a significant amount of disagreement over exactly how to characterize Lewis's notion of naturalness, and whether all of the features mentioned in the quote can be held together consistently.[3] One thing that is agreed upon is that Lewis's notion of naturalness comes in degrees, in that properties can be more or less L-natural. Thus, for example, properties can be more or less fundamental. This point is particularly relevant when it comes to the nature of truth, as few (if any) hold that truth is one of the fundamental properties of the universe. This is partly because it is generally thought that truth depends on being, in the sense that the way things are determines the truth value of sentences, and not the other way around. For example, the sentence "grass is green" is true because of some state of the world, namely grass's being green. On the contrary, grass is not green because of the truth of the sentence "grass is green": the truth value of the sentence depends on the way things are, and not vice versa. Thus, truth is not taken to be a fundamental part of reality, as whether or not a sentence is true largely depends on some other state of affairs. Accordingly, what we will be looking for in regard to fundamentality is not the claim that truth is a fundamental property of the universe, but at *how* fundamental truth is: what truth depends on, and what depends on truth.

How L-naturalness relates to T-naturalness is a complicated matter, but there is reason to think that Lewis's notion of "natural property" differs from the traditional conception, and we will treat them as separate in what follows. A property will be L-natural, then, insofar as it marks similarities between its bearers, and insofar as it is fundamental.

The Correspondence Theory of Truth

The correspondence theory of truth is perhaps historically the most influential theory of truth, and gives expression to the broadly "realist" tradition which holds that a true belief correctly represents the mind-independent world. One primary way in which correspondence theories differ is in how exactly they understand the notion of representation.

Representation as Isomorphism

Perhaps the best known form of the correspondence theory of truth is the view that truth is a matter of the isomorphism, a shared structure or form, between a belief and a worldly state of affairs, or fact.[4] True beliefs fit the facts, as it were, as a glove fits a hand, as they share some structural resemblance with the facts. On this view, beliefs are thought of as complexes, in which a certain object – this chair, for example – is represented to have a certain property – woodenness. Beliefs have a certain structure, which represents an object as having a certain property. A belief is true on this theory just in case there exists a fact – some mind-independent state of affairs – which exhibits a structure similar in form (or "isomorphic") to the belief. Using our chair example, the belief in question would be true just in case there exists a fact which is composed of the object in question – this chair – instantiating the property in question – woodenness. One analogy sometimes used to bring the idea out further is to think of the view in

3 See, for example, Schaffer (2004). For an exhaustive study of L-naturalness, see Dorr and Hawthorne (2013).

4 Classic expressions of this view are given by Russell (1912) and Wittgenstein (1921).

terms of pictorial representation: beliefs "picture" the world to be a certain way, just as a (realistic) painting depicts a landscape to be a certain way, and a belief is true just in case it correctly depicts reality. Just as a picture is an accurate representation of a scene in virtue of the relations between items in the picture matching – in a sense – the actual relations between the items depicted in reality, so a belief is true when the structure of the belief matches the structure of a worldly fact.

Despite the intuitive appeal of the isomorphic correspondence theory, many who favored the correspondence approach have turned away from it, interestingly because it seems to fare badly on naturalistic grounds. One reason for this is suspicion of one of its aspects: namely, *facts*. Facts are thought to be the neat units that make up reality, and beliefs are true or false insofar as they match the structure of these facts. However, suspicion has grown over this notion of a fact, particularly on naturalist grounds. In respect to T-naturalness, for example, facts, as metaphysical units of reality, do not seem open to empirical investigation, even if their component parts are. For example, we may perceive a chair, and woodenness of the chair, but it is not clear that we also perceive *the fact* that the chair is wooden. Moreover, there are also concerns about facts themselves having any causal powers over and above their component parts. Again, the woodenness of the chair may cause the chair to burn, but it is not clear that *the fact* that the chair is wooden causes anything. As a result, many theories of representation reject facts as *metaphysical* entities, thinking that they at the very least oversimplify things, and at the most are unnecessary and naturalistically problematic.

Representation and Reference

As a consequence, many who favor a correspondence view of truth have turned to a different proposal, which understands representation in terms of *reference*, or denotation. The basic idea is that a simple subject-predicate sentence is true just when the object referred to by the singular term has the property referred to by the predicate term. This is an idea developed by two influential modern theories of truth, namely those of Devitt (1984) and Field (1972).

To explore this thought further, let us look at what Lynch (2009) takes to be the essence of such theories of truth:

REPRESENT: The belief that a is F is true if and only if the object denoted by <a> has the property denoted by <F> (2009, 24).

REPRESENT is different from isomorphic correspondence in that there is no commitment to a structural match between a belief and a fact; all that is at issue is the reference, or denotation, of the concepts involved in the belief. Lynch notes that the additional thing that a theory needs to add is a satisfactory theory of reference, giving causal and teleological theories as two examples:

CAUSAL: <cat> denotes cats = cats, under appropriate conditions, cause mental tokenings of <cat> (2009, 25).

TELEOLOGICAL: <cat> denotes cats = the biological function of <cat> is to be mentally tokened in presence of cats (2009, 26).

These toy theories of reference are then plugged into REPRESENT to give two models of truth:

CC (Causal-Correspondence): The belief that a is F is true if and only if the object causally mapped by <a> has the property causally mapped by <F> (2009, 27).

TC (Teleological-Correspondence): The belief that a is F is true if and only if the object functionally mapped by <a> has the property functionally mapped by <F> (2009, 27).

Finding a satisfactory theory of reference has proved the main obstacle for theories of this type, but, as we have seen with the causal and teleological models, there are options. For the purposes of this chapter, we needn't take a stand on *which* theory of reference is the correct one, and what we say can apply equally to causal and to teleological conceptions.

Representational Reference and T-Naturalness

There are some grounds to think of a theory that understands truth in these terms as reasonably T-natural. The key role that representation of this kind plays in cognitive science suggests that there are good prospects for this. Indeed, as Lynch (2009, 22) notes, representation of this form is a key part of current theories in cognitive science that attempt to understand human cognition.[5] In particular, cognitive scientists are interested in a distinction that is important as regards truth: the distinction between correct and incorrect representations.[6] The view of truth under discussion in this section, which understands representation in terms of reference, allows us to say something more specific by virtue of the role of reference. Reference, just as with representation, is a relation between mental or linguistic entities and the world that they encounter. Accordingly, it is not unreasonable to suppose that it is a relation in principle open to the kinds of empirical investigation that cognitive scientists pursue. Indeed, the CAUSAL and TELEOLOGICAL theories seem to suggest an openness to empirical investigation through the notions of causation and biological functioning, respectively. In terms of having a causal role, both of these theories of reference do well. For example, the CAUSAL theory directly uses the notion of causation, and, if truth is understood in terms of a causal theory of reference, this gives truth a direct place in the causal order. In terms of the TELEOLOGICAL theory, even though reference is not understood directly in terms of causation, there is good reason to think that the key notion of biological function has a causal role to play. Thus, if truth were understood in terms of a theory of reference based on biological function, then truth would inherit a particular causal role.

Representational Reference and L-Naturalness

What does the referential theory offer in terms of grounding similarities? At the very least, it seems to give us a common standard that all true beliefs have to meet, encapsulated by the principle REPRESENT. Further similarities can be brought out depending on which theory of reference we plug into REPRESENT. For instance, if we plug in the CAUSAL theory, then all

5 For an example of a neuroscientific account of representation, see, for instance, Ryder (2004).

6 For other philosophical views along these lines, see, for instance, Fodor (1987), Millikan (1984), and Papineau (1987).

true beliefs will have the proper causal relations between their components and the objects and properties they are about. If we plug in the TELEOLOGICAL theory, then all true beliefs will have proper functioning of their component parts. As with isomorphic correspondence, there are limitations to the kinds of similarity that can be grounded here, as different beliefs will be made up of different concepts, which will denote different objects and properties, but we should at least get some general standard that all true beliefs must meet in order to be true beliefs.

In regard to fundamentality, we must turn to how fundamental reference is thought to be in order to understand how fundamental truth would be on the referential account. As with truth, reference is not plausibly one of the universe's fundamental features, particularly as a lot needs to be in place before reference is on the scene, namely an item referring (a sentence, say, or a statement) and an item to be referred to (say, an object in the world or a state of affairs), both of which would be more fundamental than the relation of reference between them. However, this is not to say that philosophers have not thought about reference as being reasonably fundamental, even if in a fairly restricted way. For instance, Williamson has recently suggested that we should think of the relation of reference as "doubtless not a *perfectly* natural one, for the most basic structure of the world is not mental, but natural by the standards of mentality" (Williamson 2007, 268). If this idea can be defended satisfactorily, then the referential account of truth will be able to claim that truth is fundamental to the degree expected of mental or semantic properties.

Epistemic Theories of Truth

Epistemic theories emphasize the intimate link between truth and ordinary human practices of assessment and inquiry. In this respect, rather than being a matter of agreement or structural resemblance with a mind-independent reality, truth is generally thought of as a function of idealized justification, and, as such, it is directly connected with our beliefs at some idealized stage of inquiry.[7]

The main inspiration for such a conception of truth comes from the American pragmatists, in particular from C.S. Peirce. According to Peirce (1878), "The opinion which is fated to be ultimately agreed to by all who investigate is what we mean by truth." The various versions of the epistemic conception differ primarily with respect to how they conceive of the idealization involved in the characterization of truth: as involving the idea of an ultimate stage or limit of human inquiry (Peirce 1878), or the notion of ideal (or good enough) epistemic circumstances (Putnam 1990), or the notion of justification that remains undefeated in any advancement of inquiry (Wright 2001).[8]

7 Though Rorty's attitude toward truth and its significance in philosophical theorizing is rather complex, one might see Rorty as an advocate of a conception of truth broadly in line with the epistemic conception, without, however, trafficking with the kinds of epistemic idealization typically associated with it. Of course, this shouldn't be taken, in the case of Rorty, as an attempt to provide an analysis of truth in terms of some nonidealized epistemic notion (e.g., justification), since Rorty agrees with Davidson that no theory of the nature of truth is possible. The thought is that the only criterion we have for applying the word "true" is justification. However, as Rorty writes, "the question 'Do our practices of justification lead to truth?' is as unanswerable as it is unpragmatic. It is unanswerable because there is no way to privilege our current purposes and interests. It is unpragmatic because the answer to it would make no difference whatever to our practice" (Rorty 1998, 4).

8 See also Goodman (1978, 123–124). For a criticism of both Peirce and Putnam's epistemic conceptions of truth, see Wright (1992, 37–42; 2001, 759–770).

The account of truth based on the last kind of idealization – what Wright calls *superassertibility* – represents a significant improvement on both Peirce's and Putnam's' accounts, and, for this reason, is taken by many to be the most promising version of the epistemic conception of truth. The thought is that, at least in certain domains of discourse, truth coincides with superassertibility,[9] where superassertibility is defined as "the property of being justified by some (in principle accessible) state of information and then *remaining* justified no matter how that state of information might be enlarged upon or improved" (Wright 2001, 771).

Since every version of the epistemic conception of truth is grounded, in one way or another, on the notion of justification, the question of whether epistemic truth qualifies as T-natural or L-natural ultimately depends on the question of whether the underlying notion of justification that these theories deploy qualifies as such. As a general roadmap, we can observe that if we ground epistemic truth on a broadly externalist conception of justification[10] – for example, various forms of reliabilism[11] – then the chances are that we will end up with a notion of truth that qualifies as reasonably natural. After all, the question of whether a certain cognitive or perceptual system is working reliably – that is, whether it reliably tracks an external world of objects – is open to empirical investigation, and a belief's being grounded in a reliable cognitive or perceptual process seems to be the right kind of thing to have causal efficacy. This suggests some promise for T-naturalness. In terms of L-naturalness, possessing the (truth) property of *superreliability* makes for genuine similarities among its bearers: they all share the property of being formed on the basis of an enduringly reliable cognitive or perceptual mechanism. Lastly, reliability might turn out to be a reasonably fundamental feature of cognitive and perceptual systems.

On the other hand, if we opt for an internalist account of justification – for example, some form of coherentism[12] – then the chances are, depending on how exactly the notion of coherence is cashed out, that our notion of truth (e.g., supercoherence)[13] will qualify as natural to a much lesser degree than its externalist cousin. For example, considering

9 Wright does not subscribe to the view that truth is superassertibility *everywhere*, but rather to the view that superassertibility is the right notion of truth for those domains of discourse where the idea of truths that outsource our cognitive and epistemic capacities seems inappropriate (e.g., moral and comedic discourse). In fact, Wright advocates a *pluralistic* conception of truth.

10 Internalism and externalism, as they are understood in this chapter, are rival conceptions of epistemic justification. Though there is a flourishing and ongoing debate about what exactly these two conceptions amount to, for the purposes of our discussion the following rough characterization should suffice. Internalists generally hold that a thinker can have some form of access to the epistemic credentials of her belief. In other words, the core idea is that the thinker can be aware of the basis for the justification of her belief. Externalists, by contrast, deny that this kind of accessibility must be a necessary feature of epistemic justification. They generally claim that it suffices for a thinker to employ a method of inquiry that meets some external criteria of reliability in order to be justified in holding a certain belief. The important point for us is that whether a method of inquiry is reliable or not (and thus, whether a certain belief is epistemically justified or not) is, for an externalist, an empirical matter that can be ascertained by means of a purely naturalistic methodology. For an introduction to internalism and externalism in epistemology, see BonJour and Sosa (2003).

11 In very general terms, reliabilism in epistemology indicates that group of theories of epistemic justification that insist on the truth-conduciveness of a belief-forming process or method. For an accessible introduction to various forms of reliabilism in epistemology, see Goldman (2011).

12 Coherentism about justification is, roughly, the thought that justification is structured like a web, where the epistemic strength of any given belief depends on the strengths of the surrounding beliefs. For an introduction to various forms of epistemic coherentism, see Elgin (2005).

13 See Capps, Lynch, and Massey (2009) for a relativistic interpretation of the notion of supercoherence.

T-naturalness, the coherence of a belief with other beliefs is not so much an empirical matter as a logical and epistemic matter: we do not need to carry out empirical investigations in order to discover whether a system of beliefs coheres; rather, we need to look at the evidential and logical relations among the beliefs. This also throws some doubt on whether coherence fits into the general network of causal interactions, as the matter of coherence itself does not seem to be determined by causal relations. For L-naturalness, there is an established complaint[14] that the notion of coherence itself relies on an independent notion of truth. The thought is that we want a theory that it is able to discriminate between good cases of coherent sets of beliefs (i.e., those that are in agreement with reality) and bad cases (i.e., those that do not agree with reality). Because it is possible to have maximally coherent sets of false beliefs, in order to discriminate between good and bad cases, an understanding of truth in terms of agreement with reality must already been in place. This fact will render this understanding of justification less fundamental than whatever other notion of truth needs to be presupposed. There will, however, be some prospects for genuine similarity, as all true beliefs will share the property of cohering with the relevant system of beliefs.

Deflationism

Both correspondence and epistemic theories assume that truth has a substantive nature; it is a property that calls out for a robust metaphysical theory. In contrast, so-called "deflationary" views about truth take this assumption to be mistaken – they "deflate" the metaphysical pretensions of their more traditional rivals. The deflationist holds that you are not really saying anything more when you say "'grass is green' is true" than when you say "grass is green." In this respect, the truth predicate, "is true," has no metaphysical role to play; that is, it adds nothing to the world.

We take "deflationism" to name a family of views distinguished by three main theses. First, all deflationists maintain that the truth predicate has an important logical function. It is a device of endorsement of a certain claim or collection of claims. If we want to endorse the claim made by Tom that Russell is a great philosopher, we can assent to it or just assert it ourselves. With respect to such ordinary endorsements, the truth predicate is not needed. However, it *is* required in cases in which endorsement is not of this ordinary and direct kind. For instance, endorsement can sometimes be indefinite: if we want to express our endorsement of some of the claims made by Tom without specifying which ones, we need the truth predicate that allows us to say, "Something Tom said is true." Also, endorsement can sometime be generalized: if we believe that Plato is the best philosopher of all time and endorse all of his philosophical doctrines, we need a truth predicate that allows us to say things like, "Everything Plato said is true." Moreover, a truth predicate is needed because we may want to endorse a certain statement by name; for instance, when we say "Gödel's second incompleteness theorem is true." Thus, deflationists claim that the truth predicate is needed to fulfill these important functions, and, in this respect, we cannot dispense of it.

The second thesis of deflationism is that "is true" does not designate a substantive property; that is, a property with an underlying explanatory nature. On the contrary, the only function that truth has is entirely expressive, not explanatory.

14 See, for example, Russell (1912) and Walker (1989).

The third feature of deflationism amounts to the thesis that all that is required for "true" to play its logical and expressive role is that it satisfies the following schema:

EQUIVALENCE SCHEMA: The proposition that *p* is true if and only if *p*.[15]

Instances of the equivalence schema are epistemologically fundamental in that we do not seek to justify them on the basis of anything more basic. As competent language users, we are disposed to accept any instance of the schema.[16] Moreover, the given equivalences are taken to be explanatorily basic in the sense that our truth talk is fully explained by them. In these respects, the various equivalences are taken as axioms of the theory of truth. The deflationist's project, then, is that of providing a theory of truth that amounts to nothing more than the totality of propositions expressed by (nonpathological) instances of the equivalence schema, which are taken as axioms of the theory.

Having specified the target of our discussion, the question is whether deflationary truth qualifies as natural under either of the senses of naturalness we are working with. At first glance, it would appear that the answer should be no, given that deflationists explicitly deny that truth has an underlying nature. However, a more careful analysis shows that things may be less straightforward than they appear to be. In particular, given that we are working with a gradable notion of naturalness, the question we should ask is to what extent, if any, deflationary truth counts as (T- or L-) natural.

Deflationary Truth and T-Naturalness

As we have seen, the question of whether a property counts as reasonably T-natural depends on whether it is open to empirical investigation and has a certain degree of causal power. Concerning causal power, deflationists emphatically deny that truth, *qua* logical property, has any causal power. Horwich, for instance, writes:

> According to minimalism, we should…beware of assimilating *being true* to such properties as *being turquoise, being a tree*, or *being made of tin*. Otherwise we will find ourselves looking for its constitutive structure, its causal behaviour, and its typical manifestations – features peculiar to what I am calling "*complex*" or "*naturalistic* properties." (Horwich 1998, 38)

This should not be taken as the claim that truth cannot properly figure in causal explanations. In fact, we do make frequent use of the truth predicate in the explanation of an array of phenomena, such as the empirical success of our true scientific theories and our success in navigating the world. For instance, we believe that certain advanced technologies – such as the GPS – work very well and reliably because of the truth of the theories on which they are based. Moreover, we believe that the truth of our beliefs, together with the presence of certain desires, explains why we are generally successful in our actions; that is, believing what is true generally gets us what we want. However, Horwich argues that truth's occurring

15 I follow Horwich here in understanding the biconditional in the equivalence schema as a *material* biconditional. Typically, the schema is formulated in the following way: <p> is true if, and only if, *p*. Here, surrounding an expression, p, by angled brackets yields an expression that refers to the propositional constituent expressed by what p says. So, <grass is green> (within angled brackets) is to be read as the proposition that grass is green.

16 Or, at least, all those instances that we are in a position to grasp cognitively, and all those that involve propositions which are not manifestly paradoxical.

in such explanations does not pose a serious threat to deflationism. On the contrary, it is just another way of illustrating the generalization function of the truth predicate that is at the core of the deflationary conception of truth.

Focusing on the issue of practical success, deflationists claim that we can give a truth-free explanation of each particular instance of practical success. For example, suppose that I want to drink some milk. And suppose that I believe that there is milk in the fridge when in fact there is milk in the fridge, and I also believe that if I go to the fridge I will have some milk. Then the chances are that, acting on the basis of my beliefs, I will get some milk and thus I will get what I want. This, according to Horwich, is a perfectly effective explanation of my practical success in satisfying my desire for milk that does not involve truth. We can thus dispense of truth in the explanation of practical success. The general principle that having true beliefs leads to practical success can then be derived by exploiting the generalization function of truth, starting from a big conjunction of the (potentially infinitely many) instances of practical success like the one described in the milk example.[17]

There might be different reasons to be unsatisfied with Horwich's proposal. One might wonder whether deflationism can actually account for the use of truth as a device of generalization in explanatory contexts (Gupta 1993, 359–369). Alternatively, one might question the explanatory efficacy of Horwich's account, arguing that it has the explanatory direction the wrong way around: it is the general principle that explains why the instances hold, and not the other way around (Lynch 2009, ch. 6). If these criticisms are on the right track, the result will be a refutation of deflationism more than a proof that deflationary truth counts as T-natural.

Finally, one might accept Horwich's account but insist that, though it shows that the truth property is not causally *efficacious*, it nevertheless fails to show that truth is not causally *relevant*. This last strategy is defended by Damnjanovic (2005), who claims that truth is not explanatorily epiphenomenal, in that it provides information that is not conveyed by an account purely in terms of causally efficacious properties.[18] However, even conceding that Damnjanovic is right in claiming that truth is a causally relevant property,[19] it does not follow without some further argument that truth qualifies as a T-natural property to a significant extent. It may well be that, though truth being causally relevant makes it more T-natural than a purely T-unnatural property, the extent to which truth qualifies as T-natural is minimal and thus compatible with the deflationary conception.

17 See Horwich (1998, 22–23, 44–45) for a formal illustration of how this works in details. See also Damnjanovic (2005, 56–57).

18 See Lynch (2009, 120–127).

19 Horwich might insist that the case of truth is one of those in which the causal relevance of the property is merely epiphenomenal, in that it does not provide any further information to that conveyed by the underlying causally efficacious properties. In this respect, truth would not essentially occur in an explanatorily adequate description of the causal process that brings about practical success. To illustrate the point, we can use the following example: "A fragile glass is struck and breaks. Why did it break? First answer: because of its fragility. Second answer: because of the particular molecular structure of the glass" (Jackson and Pettit 1990). The property of being fragile is causally relevant without being efficacious, while the particular molecular structure of the glass is also causally efficacious. However, Jackson and Petite note that "under natural background assumptions, we can say that someone who understands the molecular structure which accounts for the breaking of the glass understands all that is grasped by someone who offers the fragility account" (1990). In this respect, the fragility account is explanatorily epiphenomenal. Horwich can claim that the explanatory relevance of truth in practical success is similar to that of the fragility of the glass in this example.

With this in hand, what should we say about the question of whether deflationary truth is open to empirical investigation? An answer to this question will depend in part on what deflationists can say about the previous mark of T-naturalness. If they can show that truth is neither causally relevant nor efficacious, then deflationary truth, *qua* logicosemantic property causally disconnected from the empirical world, will not be open to empirical investigation. If, on the other hand, deflationists are compelled to concede that truth is a causally relevant property, and that it is so in a not-merely-epiphenomenal sense, they will be committed to maintaining that, at least with respect to that feature, deflationary truth is open to empirical investigation. After all, the question of whether a certain property is causally relevant with respect to a given phenomenon is something that can be addressed by empirical methods. If this is the case, the very coherence of deflationism will be under serious pressure.[20]

Deflationary Truth and L-Naturalness

As we have seen, the extent to which a property qualifies as L-natural depends on two main factors: whether it grounds genuine similarities between bearers and whether it qualifies as fundamental.

Concerning the fundamentality feature, it is safe to assume that deflationary truth is not one of the fundamental features of the universe. By definition, deflationary truth has no underlying nature, and as such it has no place in the fundamental ontology postulated by our best scientific theory of the universe. We can, of course, make use of the generalization function of the truth predicate in formulating our best scientific theory, but no ontological commitment at the fundamental level will follow from such a use.

This granted, it might still be the case that truth qualifies as reasonably fundamental in a purely logicosemantic sense. In fact, though deflationists typically maintain that there are things to say about truth – such as that truth plays an important logical function and that everything there is to say about truth can be captured by means of the equivalence schema – they nevertheless believe that truth is neither metaphysically reducible to nor conceptually analyzable in terms of more fundamental semantic notions, such as reference and satisfaction. This means that in providing a scientific theory of language and thought, truth will feature among the fundamental notions of such a theory. In this respect, deflationary truth, though nonfundamental under a general conception of fundamentality, does not qualify as a "gerrymandered" property.

What does the deflationary theory have to say about the grounding similarities feature of L-naturalness? A variety of entities can play the role of the bearer of deflationary truth. Because deflationists maintain that truth has no underlying nature, they deny that true propositions share some additional common feature just in virtue of their possessing the truth property. However, deflationists are also generally happy to concede that truth represents a mark of correctness, in the sense that true propositions are those that it is correct to believe (see, e.g., Field 1994, 265; Horwich 1998, 139). They take this to be an important aspect of truth. But they also insist that such an aspect can be fully accounted for by the deflationary conception and is thus compatible with the thesis that truth has no underlying nature (see

20 Whether this case can be made compatible with a broadly deflationary conception of truth is a question that requires a much more extensive discussion than we can pursue in the limited space of this chapter. But see Damnjanovic (2005) for a negative answer.

Horwich 1998, 20–23, 139–144). The question of whether (and to what extent) deflationists are right in claiming that they can account for this normative aspect of truth is a matter of controversy that we cannot fully explore here (see Dummett 1959; Wright 1992; Lynch 2004; 2009; Ferrari 2014). The point we would like to emphasize is that if this normative aspect of truth can be made compatible with a broadly deflationary conception, then it seems as if there is some feature other than truth that all and only true propositions have in common – namely, being a proposition that it is correct to believe. Even conceding that this feature is a rather minimal one, in the sense that it requires no substantive explanatory work on the part of truth in order to be accounted for, it might still be enough to grant to truth a nontrivial role in marking a genuine similarity between its bearers.

In conclusion, what our discussion has shown so far is that though deflationary truth does not qualify as significantly L-natural, it might nevertheless possess a positive, though rather small, degree of L-naturalness.

Truth Pluralism and Naturalism

A fourth kind of truth theory is called pluralism. In contrast to monists, who hold that there is only one interesting property of propositions in virtue of which they are true, and deflationists, who can be understood, crudely, to deny that there is any such interesting property, the pluralist seeks to widen the playing field.[21] She endorses:

PLURALISM: There is more than one property of propositions in virtue of which propositions are true.

While there are, as we'll see, a range of pluralist positions on truth, the basic idea behind most versions is that while some propositions are true in virtue of, say, corresponding to reality, others may be true by virtue of possessing some epistemic property, such as superassertibility. Put differently, representational theories of truth might be right for some areas of our thought and talk, and pragmatist or epistemic views might be right for others.

Pluralist theories can themselves be divided into two rough groups. So-called "strong" pluralist theories hold that there are literally different properties denoted by the truth predicate. When ascribed to certain kinds of content, "true" picks out one property (should that content be true), but when ascribed to a different kind of content, it picks out another. In this way, the word "true" is referentially ambiguous, and there is no property that all and only true propositions share in common. "Moderate" truth pluralists, on the other hand, hold that there is such a property, that the truth predicate always refers to that property, but that this one property can be multiply realized or manifested by distinct properties for different kinds of content. Truth, on this view, is a functional, multiply realized property. Sometimes it is realized by a representational property, and sometimes by superwarrant.

The extent to which pluralist theories are consistent with naturalism is at best unclear.[22] Obviously, it depends on which kind of truth pluralism and which kind of naturalism

21 Crispin Wright is the most important advocate of pluralism. His original statement is Wright (1992), and some important revisions to his view were made in Wright (2001); there, as here, the position is presented in terms of properties.

22 Though, see Edwards (2013) for some discussion of this issue.

are in question. Strong pluralists, for example, deny that there is a general property of truth: for them, there isn't even a question of whether truth as such is natural; the issue is really only whether various properties that the word "true" ambiguously picks out – properties like superwarrant and correspondence – are natural. And the question of whether truth is natural according to moderate pluralists is no clearer. Consider whether moderate truth pluralists – who take truth as a functional property – can hold that truth is T-natural. That question appears to hinge on whether one takes functional properties in general to be T-natural. Likewise, whether truth is T-natural according to strong pluralism also depends on other factors: if one takes the properties to which truth is identified (or to which the truth predicate variably refers) as T-natural, then truth will be T-natural as well.

In sum, there is little one can say in general about the relation between truth pluralism and naturalism. Whether a truth pluralist sees truth as either T- or L-natural will depend on her other views, both about properties like superwarrant and correspondence and about the metaphysics of properties in general.

Finally, it is worth remarking that some pluralists – like Lynch (2009) – take it to be a conceptual fact about truth that it is a normative property. According to Lynch, it is part of truth's functional role that it acts as a norm of belief and a goal of inquiry. Whether this precludes truth being a natural property as well depends on issues we address in the next section.

Truth as Normative

One issue that has received significant attention in recent years is the question of whether, and to what extent, truth is a normative property. A property is normative when it is a valuable property to possess. If normative properties are non-natural, then, if truth is normative, it is also non-natural. These questions are complex, and an extensive discussion of them would range far beyond the scope of the present chapter. In this section, we briefly discuss how the issue of truth's normativity might impact the four major theories about the nature of truth.

An immediate complication we can't avoid is that to say that the claim that truth is normative can mean one of two things. First, we might mean that it is normative of the act of believing. Thus, we might ask in virtue of what beliefs are in good standing, or correct, and we might hold that a belief *qua* belief is correct just when its content is true. To endorse this thought is to endorse that, in one sense of the word "norm," truth is the basic norm of belief.[23]

The second thing that might be meant by saying that truth is normative is that the state of affairs of *having* true beliefs is a good. If so, then we should strive to realize that state of affairs while engaging in any practice aimed at producing beliefs. Inquiry is one such practice. So, having true beliefs, as it were, might be thought to be a worthy goal of inquiry: the practice of figuring out what to believe. This is a thought that again is often identified with certain pragmatists, who in some moods go so far as to identify truth with that toward which we strive during inquiry.

23 It needn't be the only norm, of course, or the only way in which a state of belief can be correct. Beliefs can also be correct when they are justified, for example, and because of that we presumably sometimes correctly believe what is false. Thus, the correctness of believing what is true must be *pro tanto*: it is always correct, but not always correct all things considered. In this way, cognitive norms like justification and truth are no different than most norms or values. Keeping one's promises is always *pro tanto* right, but it is not always right all things considered, as everyone knows. Sometimes a *pro tanto* value is outweighed by other values, and so it is with truth.

These two ways of understanding the normativity of truth sound alike, but they are actually quite different. Consider this: you might think that being a winning move is what makes a particular move in a game correct but wonder why we should care about winning. In the same way, you might think that being true is what makes a belief correct but wonder why we should care about having correct beliefs. You might wonder why having true beliefs is valuable.

Here are two different thoughts that folks have in mind when they talk about the value of truth:

NORM: Beliefs are (*pro tanto*) correct just when their contents are true.
GOAL: Having true beliefs is of value, and therefore should be the goal of inquiry.

There is, of course, much debate – and there should be – about how to exactly characterize either of these values. But for our purposes, we can put most of these complications aside, and simply ask about how NORM and GOAL imply that truth is a normative property. In the first instance, they tell us about true beliefs, or, to speak more precisely, about beliefs in true propositions. If having true beliefs is valuable, then GOAL tells us that true propositions are those we should aim at believing during inquiry. NORM tells us that true propositions are those that are correct to believe. Now, while these truisms are explicitly about true propositions, they implicitly describe the property in virtue of which propositions are true. That is, we can take them to describe the property that plays the *truth role*. That is:

N1: If a property plays the truth role, then it is had by propositions that are correctly believed.
G1: If a property plays the truth role, then it is had by propositions we ought to aim, during inquiry, to believe.

To the extent to which N1 and G1 tell us about the property that plays the truth role, they also tell us about that role itself: its shape and structure, as it were.

Of course, it is an open question how "deep" the information they give us about the truth role really is. Thus, one might take it that what N1 and G1 tell us is only that truth is instrumentally valuable. A property is instrumentally valuable if something's possessing that property is valuable as a means to something else that is valuable. A bill of currency's being worth a dollar is instrumentally valuable. The note's having that property is valuable, but only because one can then use the note to purchase something else.

For all we've said so far, all of the theories that we've discussed in this essay could agree that truth is instrumentally valuable in either or both the sense of N1 and of G1. Thus, every theory could agree that, for example, having true beliefs is a valuable goal – one we ought to pursue – because having true beliefs helps to get us what we want. If I want a drink of water, then having a true belief that there is water in my cup will help me to satisfy that want. If I want to avoid being eaten by a rampaging lion, then having the true belief that there is a shelter nearby that can help me escape the lion will help satisfy that want. And so on.[24]

Matters get more complicated, however, when we ask whether there might be more than instrumental value at stake here. In other words, could believing what is true be valuable in and of itself – or intrinsically valuable? And if so, would this fact – again, if it is a fact – be

24 Indeed, one might go further and say that one reason that it is *pro tanto* correct for me to believe any particular true proposition (as NORM and N1 suggest) is that doing so is underwritten by the instrumental value of having true beliefs (the instrumental value of achieving GOAL).

consistent with holding that truth is T-natural? Suppose we take correspondence as our example, where correspondence is understood as a kind of, or supervening on, T-natural properties. Could a proposition's corresponding to some bit of the world make it good to believe, whether or not doing so brings any benefit to the believer?

Many philosophers – including Lynch (2004) – have thought not. The basic idea is that properties the having of which is more than instrumentally valuable could not be merely T-natural properties. But the issues here are complex. One source of complexity is yet another distinction: that between something being valuable in and of itself and something being *a kind of intrinsic value*. Being pleasurable is arguably a T-natural property. It is also arguably valuable in and of itself: we might pursue it for its own sake, even if its value can be overridden by other values. But a separate question is whether pleasure is a distinct kind of value – a distinct species of the good. Arguably, it is – and one might think that is part of what explains why it is intrinsically valuable. So, if truth, like pleasure, or love, is a species of the good, and if it is, on that account, intrinsically valuable, then whatever property we identify with truth (like corresponding to reality) is also going to be intrinsically valuable. And if correspondence is understood as a T-natural property, then perhaps T-natural properties can have intrinsic value after all.

Another source of complexity is that a property can have more than instrumental value without being intrinsically valuable. Some things and properties can have what is sometimes called "final value" – they are worth caring about for their own sake, for example – because they are constitutive elements in something that is intrinsically valuable. Lynch (2004) argues that truth has this sort of value, in that the goal of having true beliefs is a constitutive part of elements of human flourishing. Ferrari (2014) argues that this view is consistent with deflationism. If the comparison with pleasure is on the right track, it may also be consistent with the idea that truth, or at least some of its realizers, are T-natural.

Conclusion

The question, "What is truth?" is famously vexing – even for philosophers. In the light of our reflections in this chapter, one may think that the question that we've focused on here, "Is truth a natural property?", admits of no straightforward answer. Perhaps the right thing to say is that no matter what the nature of truth turns out to be, it has both natural and non-natural features in both senses of the word "natural" previously discussed. To the extent that truth – or a property that plays the truth role – involves a causal/representational relation to the world, truth is to some extent a natural property. But truth, or the properties that realize it, may have a normative role to play as well – and to that extent be less natural in one sense. This should not be surprising. Truth – perhaps like the notions of representation, existence, property, content, and the like – is so entrenched in our conceptual framework that it resists easy categorization and definition alike.

References

BonJour, L. and Sosa, E. (2003). *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Oxford: Blackwell.

Capps, D., Lynch, M.P., and Massey, D. (2009). A Coherent Moral Relativism. *Synthese* 166(2): 413–430.

Damnjanovic, N. (2005). Deflationism and the Success Argument. *Philosophical Quarterly* 55(218): 53–67.

Devitt, M. (1984). *Realism and Truth*. Oxford: Basil Blackwell.

Dorr, C. and Hawthorne, J. (2013). Naturalness. In *Oxford Studies in Metaphysics*, vol. 8, edited by K. Bennet and D. Zimmerman, pp. 3–77. Oxford: Oxford University Press.

Dummett, M. (1959). Truth. *Proceedings of the Aristotelian Society* 59: 141–162.

Edwards, D. (2013). Naturalness, Representation, and the Metaphysics of Truth. *European Journal of Philosophy* 21(3): 384–401.

Elgin, C. (2005). Non-Foundationalist Epistemology: Holism, Coherence and Tenability. In *Contemporary Debates in Epistemology*, edited by M. Steup and E. Sosa, pp. 156–167. Malden: Blackwell.

Ferrari, F. (2014). Disagreement and the Normativity of Truth beneath Cognitive Command. PhD thesis, University of Aberdeen.

Field, H. (1972). Tarski's Theory of Truth. *Journal of Philosophy* 64(13): 347–375.

Field, H. (1994). Deflationist Views of Meaning and Content. *Mind* 103(411): 249–285.

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Goldman, A. (2011). Reliabilism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/spr2011/entries/reliabilism/ (last accessed July 15, 2015).

Goodman, N. (1978). *Ways of Worldmaking*. Indianapolis: Hackett.

Gupta, A. (1993). A Critique of Deflationism. *Philosophical Topics* 21: 57–81.

Horwich, P. (1998). *Truth*, 2nd edn. Oxford: Oxford University Press.

Jackson, F. and Pettit, P. (1990). Program Explanation: A General Perspective. *Analysis* 50(2): 106–109.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61(4): 343–377.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Lynch, M.P. (2004). Minimalism and the Value of Truth. *The Philosophical Quarterly* 54(217): 497–517.

Lynch, M.P. (2009). *Truth as One and Many*. Oxford: Oxford University Press.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Papineau, D. (1987). *Reality and Representation*. Oxford: Basil Blackwell.

Peirce, C. (1878). How to Make our Ideas Clear. *Popular Science Monthly* 12: 286–302.

Putnam, H. (1990). *Realism with a Human Face*. Cambridge, MA: Harvard University Press.

Rorty, R. (1998). *Truth and Progress: Philosophical Papers 3*. Cambridge: Cambridge University Press.

Russell, B. (1912). *The Problems of Philosophy*. London: Williams and Norgate.

Ryder, D. (2004). SINBaD Neurosemantics: A Theory of Mental Representation. *Mind and Language* 19(2): 211–240.

Schaffer, J. (2004). Two Conceptions of Sparse Properties. *Pacific Philosophical Quarterly* 85: 92–102.

Walker, R.C.S. (1989). *The Coherence Theory of Truth: Realism, Anti-Realism, Idealism*. London: Routledge.

Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*, translated by Charles Kay Ogden. London: Routledge.

Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.

Wright, C. (2001). Minimalism, Deflationism, Pragmatism, Pluralism. In *The Nature of Truth*, edited by M.P. Lynch, pp. 751–787. Cambridge, MA: MIT Press.

19

Against Naturalism about Truth

BERIT BROGAARD

Naturalism

It is fairly common in philosophy to distinguish between two forms of naturalism: methodological and ontological (Devitt 1994; Loewer 1997; Papineau 2009). Within philosophy, methodological naturalism is a view about philosophical practice. It states that philosophy and science are pursuing the same ends and should use similar methods to reach those ends. If the aim of science is to increase our collective reservoir of *a posteriori* knowledge by testing theories against the empirical data, then according to methodological naturalism, this is also the aim of philosophy. I will have nothing to say about the plausibility of methodological naturalism in this paper.

In contrast with methodological naturalism, ontological naturalism concerns not philosophical practice but what there is. It will be helpful to distinguish between two forms of ontological naturalism (hereafter, just "naturalism"). I shall take naturalism, in its strongest form, to be the view that all truths (i.e., truth bearers, such as propositions and sentences, that have *the property of being true*) can be deduced, at least in principle, from truths about natural entities on the lowest level of organization; for example, truths about the elementary particles and forces of the universe.[1] So, a conjunction of all truths about the elementary particles and forces, including the laws of nature, would *a priori* entail all truths. I define naturalism, in a weaker form, as the view that the only entities that can be causally efficacious are physical entities. This view is also sometimes captured as the idea that the physical realm is causally closed; that is, all physical effects can be explained by microphysical

1 It is important to keep in mind that we are using "true" and its cognates in these two different ways: that is, to refer to a property that truth-bearers have and to refer to the truth-bearers themselves when they have that property. The question I am concerned with here is whether naturalism about the truth property (rather than the truth-bearers) is possible. Naturalism about certain truth-bearers (e.g., token sentences) certainly seems possible.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

phenomena and fundamental microphysical laws. Call these two views *strong and weak ontological naturalism*, respectively:[2]

Strong Ontological Naturalism: All truths can be deduced, at least in principle, from truths about physical entities at the lowest level of organization.
Weak Ontological Naturalism: Only physical properties can be causally efficacious.

As strong ontological naturalism rules out the possibility of any supernatural or nonphysical entities, it *a fortiori* rules out that there can be supernatural entities that can causally influence the spatiotemporal world. So, strong ontological naturalism entails weak ontological naturalism. But the weak view does not entail the strong. Unlike the strong view, the weak view is compatible with the existence of supernatural and nonphysical entities, such as ghosts, angels, and golden mountains, as long as these supernatural and nonphysical entities cannot causally influence the world. However, since physically verifying the existence of these entities would require that they be causally efficacious, for example by emitting radiation that can be physically measured, weak ontological naturalism rules out that we could ever physically verify the existence of such entities.

Strong ontological naturalism is a fairly controversial view among both philosophers and scientists. Thinkers who believe that there are strongly emergent properties, for example, will deny that all truths are deducible even in principle from truths about elementary particles and forces. Weak ontological naturalism is more widely accepted, though it is by no means close to universally accepted, even among scientists. Some consciousness researchers, for example, believe that consciousness is a fundamental nonphysical property that nonetheless may be causally efficacious (Koch 2012). Our discussion here will be limited to naturalism about truth. I will explore which theories of truth are compatible with strong and weak ontological naturalism. As we will see, the only theories that appear to be compatible with strong ontological naturalism are deflationary theories, but, as I will argue, deflationary theories are unlikely to be attractive to folks with strong naturalist inclinations.

2 I do not rule out the possibility of a moderate version of ontological naturalism that formulates naturalism in terms of metaphysical entailment rather than deducibility. For example, some physicalists deny strong ontological naturalism, as I have formulated it, but would hold that truths about the microphysical realm metaphysically entail all truths (or at least, all truths about the physical and the mental) (e.g., Levine 1983; Tye 1995; Lycan 1996; Hill 1997; Block and Stalnaker 1999). This view is sometimes cashed out as the supervenience claim that the mental supervenes on the physical. If the mental supervenes on the physical, any changes in mental properties must be accompanied by changes in physical properties. I will not discuss this more moderate form of ontological naturalism here, as none of the main accounts of truth is a candidate to be a theory that is metaphysically entailed by, but cannot be deduced from, physical theories.

There are also theorists who argue that property dualism is compatible with strong naturalism, because they take mental properties to be natural properties governed by natural psychophysical laws (e.g., Chalmers 1996). We could easily accommodate this type of view by substituting "natural" for "physical" in our formulation of strong ontological naturalism. However, in my opinion, it is ill-advised to attempt to argue for a broader definition of "natural" within the limits of a short chapter. If there are indeed primitive mental properties, then they may well turn out to be natural properties. But it doesn't follow from such a concession that all primitive properties should be counted as natural properties. If that were the case, then naturalism might quickly become a rather trivial position, as any entity that is not reducible to physical properties and that seems to have some claim to existence might then be rendered a natural property. As I don't know how to avoid trivialization on a broader formulation of naturalism, I shall here assume that naturalism requires deducibility from the microphysical arena.

Is Truth Causally Efficacious?

Weak ontological naturalism holds that only physical properties are causally efficacious. So, if truth is not causally efficacious, weak ontological naturalism can be correct even if truth is not a physical property. So, one pertinent question to ask is whether truth is causally efficacious.

It is widely held that truth has certain normative properties, such as being the aim of belief and assertion and the end of inquiry (Dummett 1959; Peirce 1992; Lynch 2009). Michael Dummett (e.g., 1959), for example, held that a person who is rationally asserting something is motivated by the aim of saying something true. Charles Sanders Peirce argued that the most important effect of truth is that it will be revealed by scientific inquiry at some future progressed state of science. As he puts it, "the opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth" (Peirce 1992, 139).

However, there is good reason to think that even if truth is normative in either of these senses, it is not normative in the strong sense of being causally efficacious. Consider first the hypothesis that truth is central to our assessment of belief. Truth may be central to our assessment of belief insofar as we predicate it of belief, but truth does not appear to guide our assessment of belief. If you say that you believe that it's raining, and I look out the window and see no water coming down from the sky, I might say that your belief is false. However, what guides my assessment of your belief is not truth, but evidence (my seeing no water coming down from the sky). My perceptual experience provides *prima facie* justification for my belief that it's not raining, and it is on the basis of this belief that I judge that your belief is false. Though I predicate truth, or the lack thereof, of your belief, it is not truth that causes me to make this assessment. What gives rise to my assessment is my perceptual experience and the belief in which it results.

Consider next the hypothesis that truth is the aim of scientific inquiry. It is certainly true that most scientists (and philosophers) will tell us that they hope to find the truth. But the property of being true does not causally influence the progress of science or scientific practice. Consider the ancient scientific question of whether the universe has always existed or came into existence at some point in the past. Astronomer Edwin Hubble eventually found a way to answer this question. He knew that light that is emitted from an object that is moving away from an observer appears redder than the light of an object that is not moving away. This is known as "redshift." By observing distant galaxies, Hubble found that their redshift increased as a linear function of their distance. So, he concluded that the universe was expanding. From this, physicists inferred that the universe was smaller in the past and therefore had not always been in existence, or at least had not always existed as it does today. The truth properties of the truths that the scientists discovered (assuming that they are true) did not cause scientific progress or practice. What caused the progress and practice were (1) a question ("Has the universe always existed?"), (2) a hypothesis about distance and light ("Light emitted from objects that move away from the observer is redshifted"), (3) observations of light emitted from distant galaxies ("The redshift of distant galaxies increases as a linear function of their distance"), and (4) inferences to the best explanation. Truth itself played no role.

Truth may be what we are searching for, it may be what we are aiming at or hoping to reach, but it does not play a causal role in forming beliefs or in guiding scientific progress or practice. Truth, it seems, is causally inefficacious. It may be thought that these considerations presuppose particular "conservative" or "conventional" theories of truth, and that truth

perhaps can be understood to play a causal role given a more radical theory of truth, such as radical pragmatism. As we will see further on, however, radical pragmatists deny that truth is a substantial property and adamantly deny that it can have any normative effect on our ordinary or scientific practices.

Could truth be causally efficacious in a different way? It is sometimes thought that causally inefficacious properties are strange, because it seems that properties must be causally efficacious in order for us to have knowledge of them. Suppose, for example, that there are causally inefficacious ghosts. Since it appears that we come to have knowledge of things as a result of these things leaving a causal mark on our cognitive system, it would seem that we could not come to have any knowledge of ghosts that cannot causally affect us. Since it is arguable that we do have some knowledge about truth, and since knowledge about a property appears to require that the property be causally efficacious, it might seem that truth must be causally efficacious. Truth may be different from causally inefficacious ghosts in this regard. A widely accepted platitude about truth is that it obeys convention T, or the T schema, which can be expressed as follows:

Convention T: ⌜φ⌝ is true iff φ.

If truth obeys convention T, then we can come to have knowledge of truth by inference. Let it be granted for argument's sake that some of our beliefs count as knowledge. Let's say that I know that I am human. From "I am human," it follows by convention T that it is true that I am human. From this, it follows by existential generalization that something is true. By property abstraction, it follows that there is a property of truth. If I have the right sorts of logic skills, I can come to know there is a truth property. But there is nothing in this process that requires the truth property itself to be causally efficacious, which is to say, the truth property itself need not have any causal influences on my cognitive system in order for me to come to have this knowledge.

If truth is causally inefficacious, then the nature of truth (whatever it is) does not counter weak ontological naturalism. Recall that weak ontological naturalism, as we defined it, is the view that only physical entities can be causally efficacious. If truth is not causally efficacious, which it probably is not, then weak ontological naturalism can be true even if truth turns out to be a nonnatural property. So, weak ontological naturalism about truth is correct (but trivially so).

Is Truth a Natural Property? The Correspondence Theory

We have established that truth is unlikely to be causally efficacious. If this is right, then the existence of a truth property is compatible with weak ontological naturalism. The question remains, however, whether truth is a natural property or not. If it's not a natural property, then strong ontological naturalism is false. This question is more difficult, as it forces us to consider the plausibility of particular theories of truth and whether they entail that truth is a natural property.

Probably the most widely accepted view of truth is a form of the correspondence theory, which takes truth to be a correspondence between a sentence or proposition and a fact (Russell 1918/1956; Wittgenstein 1921/1961). It is a bit of a mystery what this relation of correspondence is supposed to be. In simple cases, it may appear straightforward. Consider,

for example, the sentence "Dorothy is singing." This sentence consists of a noun phrase, "Dorothy," and a verb phrase, "is singing." Let us suppose that an utterance of this sentence is true. On the correspondence theory, there is then a fact that consists of a person named "Dorothy" and the property of singing. Correspondence obtains when there is an utterance of a sentence with a certain structure and a fact with the same structure, and the constituents of the sentence denote the constituents of the facts. This story about atomic truths will, of course, need to be combined with a theory of the truth conditions for molecular, or complex, truth-bearers. Truth conditions for complex nonmodal truth-bearers can be given by truth-conditional recursion clauses that can be taken to be true by stipulation. A sentence of the form "not-*p*" is true if and only if "*p*" is false; a sentence of the form "*p* and *q*" is true if and only if "*p*" is true and "*q*" is true; a sentence of the form "*p* or *q*" is true if and only if "*p*" is true or "*q*" is true; and so on.

One common critique of the correspondence theory is that it is not clear that the world comes divided up into sentence-like facts (Austin 1950, 155; Rorty 1981). The correspondence theorist may be able to get around this worry by dropping the notion of a fact from the definition of "correspondence." For example, they might say that correspondence between an utterance of a sentence and the world obtains when the noun phrase of the sentence denotes an object that instantiates the property denoted by the verb phrase. On a widely accepted view of denotation (or reference), what names and predicates denote can be understood in terms of causation and, perhaps, speaker intentions (see, e.g., Field 1972; Kripke 1980; Fodor 1990; Loewer 1997). For example, "Dorothy" denotes Dorothy when there is a causal chain leading from the speaker's use of the name "Dorothy" to the individual who was named "Dorothy." Another approach is to take truth in language to derive from the accuracy conditions of perceptual states, and then define the accuracy conditions of perceptual states in terms of covariation and epistemically optimal conditions (Stalnaker 1984; Loewer 1997). However, as we will see in the next section, it is unlikely that there is a naturalistically respectable notion of truth that appeals to epistemically optimal conditions.

There are other well-known problems for a naturalistic version of the correspondence theory. One is that it is not clear how correspondence works for normative, modal, and mathematical claims, such as "Murder is wrong," "Someone else could have written this article," and "2 + 2 = 4." Even if "2" and "4" denote entities, they probably don't denote causally efficacious entities. So, it would seem that the correspondence theory cannot be spelled out in naturalistic terms. This, however, is not a problem specifically related to the correspondence theory, but a more general problem for naturalism. A range of modal notions, such as *causality*, *disposition*, and *law*, figure in fundamental physical theories, and therefore ought to be treated as naturalistic. The same may apply to mathematical notions. It is, however, beyond the scope of this chapter to investigate whether it is possible to account for normative, modal, and mathematical facts in terms of microphysical properties, or whether these subject matters require positing new primitive properties. (See Chapters 2, 20, 21, and 25–29 for further information on some of these issues.)

There are, however, a couple of more devastating problems for the correspondence theory from a naturalistic point of view. Even if we can find a way to resolve the problems just outlined, we cannot account for all correspondence in terms of causation and speaker intentions. It is unlikely that the correspondence between all quantificational truths and material entities can be cashed out in terms of causation and speaker intentions. The strategy might work for some cases of ordinary language. For example, "There is a shortest female spy" is true just in case there is a woman who is a spy and who is shorter than all

other spies. Provided that one can be causally acquainted with the properties of being a spy, being female, and being short, we can provide a naturalistic account of the correspondence relation in this case. However, it will be much harder to provide a naturalistically respectable account of the correspondence relation between the quantificational statements of microphysical theories and the microphysical realm. It is hard to see how we could be causally acquainted with microphysical entities in the way we are with people, cats, and dogs. This is not to say that microphysical entities, such as quarks, neutrinos, photons, and bosons, are not causally efficacious, but only that the causal relationship that may obtain between us and them will be insufficient to account for the correspondence relation in naturalistic terms. The standard move made in order to account for the denotation of theoretical entities is Ramsification, which is a way of eliminating theoretical terms in favor of existentially quantified bound variables and predicates (Ramsey 1954; Lewis 1970). For example, "There is an electron in the collider" can be understood along the lines of "There is an *x* such that *x* satisfies P_1, P_2,…, P_n (where P_1–P_n are predicates of the scientific theory positing electrons, e.g., 'is an electron,' 'has a negative charge,' etc.), and *x* is in the collider." But the predicates of physical theories do not denote properties with which we are causally acquainted. So, the correspondence theory does not seem to allow for a strongly naturalistic account of truth: it requires an irreducible, ontologically primitive notion of correspondence. Note that arguing that facts are true propositions, and that correspondence is therefore identity, along the lines of Moore (1901/02) and Russell (1904), is unlikely to help. This move just substitutes an irreducible notion of truth for an irreducible notion of correspondence.

A further problem for correspondence theorists who want to grant that physical theories can be true in virtue of corresponding to the microphysical realm is that this appears to require a form of antirealism with respect to the reality that consists of medium-sized material bodies and their perceptible properties. Physicists will tell us that, on a microphysical level of reality, there are no solid objects. On a microphysical level, a chair is just an array of particles with no obvious boundary and no real shape. If we take these claims to be true, then at least some truths about medium-sized material bodies and their property instances may turn out to depend in part upon our experiences of them. Few theorists would deny the veridicality of all of our experiences of solid, medium-sized material bodies. The veridicality of perceptual experience does not require seeing into the deepest corners of reality. Our visual experiences of the colors, shapes, and textures of medium-sized material bodies are said to be veridical when the viewing conditions and neurological systems of the perceiver fall into a normal range that is such that different perceivers within this range can, roughly, agree on what they perceive. But the notion of veridicality is intimately tied to the notion of truth (Siegel 2010). When we say that our experience of a rock being solid is veridical, this is normally taken to mean that it is true that the rock in question is solid. We can account for the tactile and action-related property of solidity by defining it as the weakly emergent property of resisting penetration at a certain level of interaction. However, the veridicality of a visual experience of a rock being solid when the rock is not interacting with other medium-sized objects cannot be accounted for in this way. Rocks have the *visually* perceptible property of being solid only relative to creatures like us who experience them as such under certain conditions. Another good example is that of color: a rose has the visually perceptible property of being red only relative to creatures like us that experience it as such in particular circumstances (Cohen 2009; Chirimuuta 2015; Brogaard forthcoming). Many ordinary truths must thus be understood in terms of experience in ideal, or sufficiently

good, experiential conditions, which include an ideally placed observer. For example, where "solid" refers to a visually perceptible property that medium-sized objects can possess even when they don't interact with other medium-sized objects, "That thing is solid" is true if and only if, if circumstances were ideal for experiencing the thing in question, it would visually appear to be solid. It seems, then, that claims about correspondence between statements and reality in some cases collapse into claims about assessment in ideal circumstances. So, not only is the correspondence theory incompatible with strong naturalism, but it also appears to be committed to elements of antirealism about truth. As we will see in the next section, antirealism about truth is also incompatible with strong naturalism.

Antirealism about Truth

In his presidential address to the American Philosophical Association (APA), "How to Be an Anti-Realist", Alvin Plantinga (1982, 64–66) argues that the only sensible way to be an antirealist is to be a theist. Antirealism in this context is the epistemic analysis of truth that says:

(AR) Necessarily, a statement is true iff it would be believed by an ideally (or sufficiently) rational agent/community in ideal (or sufficiently good) epistemic circumstances.

In every world, if it is true that ideal circumstances obtain, it would be believed that ideal circumstances obtain if they were to obtain. So, in a closest world where ideal circumstances obtain, it would be believed that they obtain. So, the right-hand side of (AR) is true. It follows that it is true that ideal circumstances obtain. Since (AR) is necessary, it is necessarily true that ideal circumstances obtain. Plantinga concludes that since human existence is a contingent matter, the ideal circumstances must be those of God. If Plantinga is right, then antirealism entails a denial of strong and weak naturalism (assuming that God is nonphysical and causally efficacious).

Plantinga's formulation of antirealism involves an ambiguity, however (Rea 2000; Wright 2000; Brogaard and Salerno 2005). The ambiguity arises when we speak loosely of "ideal epistemic circumstances." The antirealist could be saying that there are circumstances ideal for the evaluation of any truth or that for each truth there is a set of circumstances (not necessarily the same set) that is ideal for the evaluation of it. These two readings provide us with subject matter-nonspecific and subject matter-specific readings of antirealism, respectively, because the latter, but not the former, avows an epistemic idealization that is specific to the subject matter of the statement.

The correct reading appears to be the subject matter-specific reading. Different circumstances are ideal, or sufficiently good, circumstances for the evaluation of different truths. For example, extremely good auditory conditions would be required for the evaluation of "That note is a C sharp," but good auditory conditions would be neither necessary nor sufficient for the evaluation of "That wall is a shade of red." The most plausible reading of (AR) is thus a subject matter-specific reading. Let "Q*p*" mean that "circumstances are ideal (or sufficiently good) for determining whether *p*." The subject matter-specific version of (AR) can then be articulated as follows:

(AR*) Necessarily, *p* is true iff, if Q*p* were to obtain, then it would be believed that *p* obtains.

Plantinga's proof is not valid on the subject matter-specific reading, because it rests on a premise that is given a subject matter-nonspecific reading. Suppose the premise is given a subject matter-specific reading instead. Then it says, for instance, "If it is true that ideal

circumstances for determining that ideal circumstances for determining that *snow is white* obtain, then it would be believed that ideal circumstances for determining that *snow is white* obtain." Since the right-hand side is true, we can infer that ideal circumstances for determining that *snow is white* obtain. This is as it should be, as we should expect "Snow is white" to come out true. But we cannot make any inferences about ideal circumstances in general.

(AR*) has its own share of problems. It rules out a truth-conditional account of meaning. On a truth-conditional theory of meaning, the meaning of "Snow is white" is equated with the condition under which it is true. Within an antirealist framework, the meaning of "Snow is white" could be understood as the ideal circumstances in which it would be believed. But (AR*) requires that ideal circumstances are specific to the subject matter of the sentence to be evaluated. This presupposes an independent account of meanings (or subject matters). So, on pain of circularity, (AR*) cannot provide a truth-conditional account of meaning. It's a good question, and not a question that I will attempt to answer here, which other theories of meaning may be compatible with the antirealist framework.

A further problem for (AR*) is that, while it avoids Plantinga's theist implication, it has an idealist implication that seems unintuitive (Brogaard and Salerno 2005). (AR*) entails that necessarily, conditions are ideal for determining whether some statement is true. Let "Q_{Qp}" mean "Conditions are ideal for assessing Q*p* for truth," and let Q*p* be the statement "Epistemic conditions are ideal for determining whether some statement is true." In this case, $Q_{Qp} \rightarrow Qp$. For, if conditions are ideal for determining whether Q*p* (i.e., Q_{Qp}), then conditions are ideal for determining whether some statement is true (i.e., Q*p*). It can be shown that this implication, together with (AR*), entails that necessarily, epistemic conditions are ideal for determining whether some statement is true. These conditions include the existence of a properly placed epistemic agent. So, necessarily, there is an epistemic agent. An antirealist must thus deny the possibility of truth in a world where there are no epistemic agents.

The conclusion that, necessarily, there are epistemic agents, and so there are no truths without epistemic agents, is not theism but a form of idealism (Brogaard and Salerno 2005). It may be thought that this consequence is not devastating, or even counterintuitive. If the bearers of truth are sentences or utterances (rather than propositions), then perhaps there are no truths without agents. The existence of sentences is mind-dependent. The conclusion is a form of idealism and is devastating to antirealism only if it implies that there cannot be "facts" in a world lacking epistemic agents. The intuition is that facts may obtain even if true sentences do not. So long as antirealism allows that there may be facts without agents, antirealism does not entail a problematic form of idealism.

The point is well taken. That there may not be truths without minds does not by itself entail idealism, since it leaves open the question of whether there can be facts without minds. But the consequence derived by Brogaard and Salerno (2005) does not allow for the possibility of facts without minds, even if sentences are treated as the bearers of truth and falsity. It was shown that in every possible world there are agents; so, antirealism implies that there are no worlds without epistemic agents. This is a stronger claim than the claim that there are no true sentences (or utterances) without agents. And it is strong enough for a substantial form of idealism. If there are no uninhabited worlds, then *a fortiori* there are no uninhabited worlds occupied by facts.

Crispin Wright (1992) has proposed an alternative formulation of antirealism, according to which a proposition is true just in case one would be warranted in believing it regardless

of how one's state of information was extended. He thus avoids the counterfactual formulation that appeared to cause trouble for Plantinga's formulation. However, arguably, Wright's proposal is just a form of the subject matter-specific reading of (AR): if a proposition is true if and only if believing it remains warranted no matter how our information is expanded, then arguably a proposition is true if and only if in close scenarios in which our information is extended and conditions are sufficiently good for assessing that proposition, an ideal agent will believe it.

The question remains whether antirealism about truth is incompatible with strong naturalism. The antirealist defines truth partially in terms of modal properties. However, as I have already mentioned, anyone with strong naturalist inclinations will need to offer a naturalistic account of modal properties. It may be doubted that this can be done in a noncircular fashion if one is an antirealist about truth. But let us suppose that it can be done, or that modal properties are respectable primitives, as their existence is implied by fundamental physical theories. Then the question remains whether antirealism, which appears to entail a form of idealism to the effect that there cannot be any truths or facts without epistemic agents, is compatible with strong naturalism. Epistemic agents, of course, can be understood as thinkers or minds; they don't need to have material bodies. Is the antirealist's theory, so understood, deducible, in principle, from truths at the lowest level of organization? The answer appears to be "No." Here is one reason to think that it is not. The theory implies that there cannot be truths, or facts, without epistemic agents. Epistemic agents thus seem more fundamental than, say, elementary particles and forces. So, if strong naturalism is taken to imply that theories of truth must follow logically from, say, the theories of particle physics, then antirealism is incompatible with strong naturalism. Of course, there may be radical reformulations of naturalism that simply require that all properties follow from fundamental properties. If the properties of agents are fundamental properties, then antirealism could perhaps be taken to be a naturalistic theory. However, on more conventional ways of cashing out "strong naturalism," antirealism seems radically antinaturalistic.

Deflationary Theories of Truth

Another type of theory of truth that I will discuss in light of naturalism is the deflationary theory. Deflationary theories of truth hold that the meaning of "true" is fully explained by the T schema and an account of the function of "true," which is primarily to serve as a device for generalization and abbreviating conjunctions and disjunctions (Horwich 1990; Field 1994).[3] Suppose John tells you and your friend, "The seminar is canceled. Professor Brown is sick. The others are thinking about going to Pi. You should come." As you part ways with John, your friend skeptically declares: "I don't think the seminar is canceled, or that Professor Brown is sick for that matter." You beg to differ and simply reply, "No, what John said is true," thereby abbreviating the longer conjunction of John's statements. This function of the truth predicate at least partially explains why we have it in our language.

3 Some distinguish between the redundancy theory and minimalism as distinct forms of deflationism, where the redundancy theory holds that there is no truth property at all, whereas minimalism holds that there is a truth property but that it is insubstantial: there is no more to it than the disquotational instances of the T schema. However, I shall ignore this distinction here, as it is of no consequence in the present context.

Furthermore, when we master the concept, we are willing to accept all instances of the T schema. For instance:

"Snow is white" is true iff snow is white.

But according to deflationists, there is no substantial truth property in the world. We might compare the truth predicate "true" to the connective expression "or." Because "or" is an expression for a truth-functional connector, its meaning is given by its truth conditions. So, "*p* or *q*" is true just when "*p*" is true, "*q*" is true, or both "*p*" and "*q*" are true. But, arguably, "or" does not denote any language-external properties. Likewise, deflationists will say that "true" does not refer to anything outside of language itself.

More radical versions of deflationism – sometimes called "neorelativism" or "radical pragmatism" (Zemach 1987) – hold that truth is "merely an expression of commendation" (Rorty 1985, 6). As Rorty puts it, "the [radical] pragmatist says that there is nothing to be said about truth save that each of us will commend as true these beliefs which he or she finds good to believe" (1985, 5). Something like this view has been exposed by a diverse range of thinkers, including Quine (1969; 1974), Rorty (1982; 1991; 1998), and Davidson (1984).

Radical pragmatism about truth is not typically classified as deflationist. In fact, it seems that postmodernists like Rorty take truth to be intersubjective agreement among the members of a community, which would make it a substantial property. In a famous quote, Rorty states that "truth is what your contemporaries let you get away with" (1979, 176). However, as his student Crispin Sartwell (2007) argued in an obituary, these provocations are best understood as rhetorical. And, in fact, there is good reason to think of radical pragmatism as a variant of deflationism. The proponents reject the notion of truth as correspondence to a language- and theory-independent reality. The very notion of correspondence is regarded as empty, because it presupposes either that reality exists independently of our theories (Rorty 1985) or that expressions have some invariant meaning and reference (Quine 1969). Moreover, radical pragmatists explicitly state that truth is not a substantial property. The predicate "true" merely plays a role as an expression of certain attitudes (Rorty 1985).

As "true" does not express a substantial property, on the deflationary account, this theory is perfectly compatible with both weak and strong ontological naturalism. A deflationary theory of truth is no more an objection to naturalism than a conventional theory of disjunction or existential quantification. In fact, if truth simply is the semantic value of a predicate in a natural language, and truths about the human mind are deducible in principle from our best physical theories, then we can be nearly certain that features about the truth predicate will also be deducible in principle from these theories. So, those who find strong naturalism irresistible may find a deflationary approach to truth attractive.

The question remains, however, whether deflationary accounts of truth can ultimately be preserved. There have been several objections to deflationary theories in the literature. One common objection is that they cannot account for the normative aspects of truths (see, e.g., Wright 1992). I have already raised an objection to the idea that truth plays a normative role, and this objection can be turned into a reply on behalf of the deflationist. The most frequently cited normative aspects of truth are (1) that truth is the aim of scientific inquiry and (2) that beliefs are correct if and only if they are true. A plausible response to the first challenge is to say that the real guiding force of scientific inquiry is a *hope* or *desire* that we will form true beliefs as a result of scientific inquiry, where a true belief can understood in terms of the derivative T schema: For any belief that *p*, that belief is correct

if and only if *p*. A plausible response to the second challenge is to say that the real guiding element in our assessment of beliefs is evidence (which consists of beliefs and perceptual experiences).

An objection to deflationism that seems to have more bite is that, on pain of circularity, the deflationist cannot provide a truth-conditional account of meaning (Jackson, Oppy, and Smith 1994). On a truth-conditional account of meaning, the meaning of a sentence is its truth conditions. For example, "Snow is white" means that snow is white just in case "Snow is white" is true if and only if snow is white. Deflationists cannot offer a truth-conditional account of meaning because the primary truth-bearers of the deflationary theory require an independent notion of meaning. The truth-bearers in the T schema "'*p*' is true iff *p*" cannot be sentences, as the T schema is only true provided that "*p*" means that *p*. So, the truth-bearers in the T schema must be either propositions or interpreted sentences. Both notions ensure that "*p*" is assigned a meaning. The proposition that snow is white, for example, can be taken to be a level of meaning that the sentences "Snow is white," "Schnee ist weiss," and "La neige est blanche" have in common. But if the T schema, which characterizes our use of the truth predicate, is formulated in terms of a notion of meaning, then on pain of circularity, the deflationist cannot provide a truth-conditional account of meaning. Though there no doubt are other plausible accounts of meaning, the challenge of providing such an account should not be taken lightly.

Thinkers, like Quine (1969; 1974), who reject that language has meaning may not appear to face the challenge of providing an alternative theory of meaning. However, Quine's theory cannot ultimately be denying that theories are associated with any form of meaning that we can grasp. If it were, the theory would be at risk of being self-refuting. Suppose that some theory implies that theories do not have any associated meaning that we can grasp. This leads to a dilemma (Zemach 1987): either we are capable of understanding the theory in question or we are not. If we are incapable of understanding the theory, then there is no reason to bother with it. If we are capable of understanding the theory, then the theory is associated with a meaning that we can grasp; but if it is associated with a meaning that we can grasp and also claims that theories are not associated with a meaning that we can grasp, then the theory is false. So, either we cannot understand the theory in question or it is false. The theory must be rejected either way.

Another common objection to deflationism is that the theory cannot explain the correspondence intuition that sentences are true in virtue of certain things obtaining in the world (Stoljar and Damnjanovic 2010). I don't think this objection has much bite if deflationism is considered in isolation of other views, as it is clear that deflationists – *qua* denying correspondence theories of truth – are not giving full credit to this intuition. However, simply denying that theories of truth must conform to this intuition becomes less plausible when deflationism is combined with strong naturalism. Those who find strong naturalism plausible will typically hold that there is some discourse that contains predicates that express natural properties, and that those natural properties exhaustively characterize what is expressed by those predicates (Dowell 2004). It follows naturally from this idea that to say that "*x* is N" (where "N" expresses a natural property) is true is to say something substantial, viz. that "*x* is N" is true in virtue of *x* instantiating N. This suggests that while deflationism is not at odds with strong naturalism, as we have formulated it here, a deflationist truth property is too thin to satisfy the naturalist's aspirations. For example, if the strong naturalist says that charm, spin, or charge is a natural property, then this claim is insubstantial, as it doesn't correspond to anything.

Conclusion: Pluralism about Truth

I have argued that truth is unlikely to be causally efficacious. As weak naturalism says that only natural properties can be causally efficacious but is quiet about any potential non-natural properties, the existence of a truth property is consistent with weak naturalism. However, it is doubtful that any realist theory that treats truth as a substantial property will turn out to be consistent with strong naturalism. The correspondence theory will be hard pressed to offer a naturalistically respectable account of the correspondence relation that supposedly obtains between physical theories and microphysical reality. Antirealism, on the other hand, appears to entail a form of idealism to the effect that there cannot be any truths or facts without epistemic agents. Reality itself requires that there are properly placed agents. This implication appears to be incompatible with strong naturalism, as we have formulated it, as it seems to require that epistemic agents, or at least thinkers or minds, are more fundamental than physical entities, such as elementary physical particles and forces. Finally, though deflationary theories of truth are compatible with strong naturalism, this is not going to be of much help to the naturalist. Deflationary theories of truth deny that there really is such a thing as truth outside of language. Rather, there are truth predicates and the semantic values of those predicates or our use of them. As truth is not a substantial property, given deflationary theories, these theories are compatible with strong naturalism, as formulated in this chapter. But these theories are not suitable adjoints to strong naturalism, as folks with strong naturalist inclinations normally want to posit the existence of substantial natural properties.

On the face of it, this does not leave us in a very good place. All of the theories of truth considered appear to have large drawbacks. The correspondence theory of truth must posit correspondence between physical theories and microphysical reality as a non-natural ontological primitive. Furthermore, claims about correspondence obtaining between ordinary statements and the medium-sized level of reality appear to collapse into counterfactual claims about how things would appear to agents in ideal, or sufficiently good, circumstances. Antirealism about truth appears to have the implication that there cannot be any truths or facts without epistemic agents, which many would consider a rather unattractive consequence, regardless of their inclinations toward naturalism. Deflationary theories of truth either are self-undermining or are faced with the difficult task of providing a non-truth-functional theory of meaning. Furthermore, even if these theories are not at odds with strong naturalism, they threaten to trivialize it. Because there is no correspondence between language and reality, the claim that deflationary truth is natural is insubstantial.

There are theories of truth that may seem to fare better in a number of respects than the three main theories discussed in this chapter. One is the pluralist theory discussed in Chapter 18. Pluralism about truth is the view that truth is a functional-role concept (Wright 1992; Lynch 2000; 2009). In different types of discourse, different properties play the truth role, which is defined by features that derive from our pretheoretic or folk conception of truth (e.g., "Truth is the aim of belief"). It is important to emphasize that pluralism is a form of role functionalism rather than realizer functionalism. Role functionalism equates the property in question with the functional characteristics, whereas realizer functionalism equates the property in question with the realizer property that satisfies the functional characteristics. For example, on a functional conception of pain, pain may be taken to be any state which produces the belief that something is wrong with the body and a desire for the state to disappear, and which tends to give rise to anxiety, moaning, whining, or

screaming. If this functional role is played by C-fiber stimulation, then a person is in pain when she undergoes C-fiber stimulation. Role functionalists, however, would identify the pain property with the functional role property itself, not the C-fiber stimulation. Likewise, pluralists about truth take truth to be identical to the functional-role property itself. This avoids a problem that would otherwise arise in accounting for semantic properties such as validity, which depend on there being a single truth property. Consider the following argument:

If animals can feel pain, then it's wrong to let them suffer.
Animals can feel pain.
Therefore, it's wrong to let them suffer.

Valid arguments are truth preserving: the premises cannot be true while the conclusion is false. But this requires that there is a single truth property that is preserved. If there are different truth properties for moral and nonmoral claims, then the argument is not valid.

I think pluralism may be among the most plausible theories of truth. But it is worth pointing out that it does not straightforwardly avoid the problems discussed her. This is because it assumes that different properties can play the truth role for different discourses. For example, it may be argued that correspondence plays the truth role for discourse about the microphysical realm, that assessment in ideal circumstances plays the truth role for truths about the level of reality that consists of medium-sized objects and their perceptible properties, and so on.[4] Pluralism by itself does not resolve the problems that face the correspondence theory and antirealism about truth. These problems require independent solutions, but since pluralism is already pluralistic, it might not be ad hoc if the pluralist were to impose restrictions on theories of truth. For example, to avoid the unintuitive consequence of antirealism about truth that agents exist necessarily, a pluralist could disallow that assessment in ideal circumstances can itself play the truth role for the assessment of whether ideal circumstances obtain. She could argue that the question of whether ideal circumstances obtain is decided by convention. Given a pluralistic framework, this move does not feel like an odd exception to antirealism about truth, but rather a natural implication of a pluralistic framework. Pluralism also allows us to provide a truth-conditional account of meaning. On pain of circularity, an antirealist theory of truth cannot provide a truth-conditional account of meaning, because the idea of an ideal circumstance that is specific to the subject matter of the sentence that is being evaluated for truth rests on a prior conception of meaning. Within a pluralistic framework, truth is not identical to assessment in ideal circumstances but to a functional role property. This seems to avoid the circularity. However, pluralism still leaves us with the consequence that there is a correspondence property that sometimes plays the truth role and that cannot be fully reduced to microphysical properties. So, it looks like the pluralist will need to reject strong naturalism.

This, however, may be exactly as it should be. A fully reductionist theory of all semantic relations appears to run into trouble of an epistemic kind (Putnam 1983, 290–296; Zemach 1987). Suppose a given theory implies that "reference" *refers* to the natural property R. As the theory logically depends on a notion of reference, it can only be assessed for correctness or plausibility on the assumption that we have some understanding of what "reference" is,

4 This is just an example of how to accommodate some of the observations made earlier in the chapter. Wright and Lynch do not divide up the labor in this way.

independently of what the theory says that it is. This is not to say that there could not be R properties, but only that we could never confirm that they function in the way that the theory says that they do. Moreover, the theory would not be deducible from theories about the microphysical realm. So, a naturalistic theory of reference would be antinaturalistic. This strongly suggests that semantic notions are irreducible, primitive relations and hence are not natural properties of the sort strong naturalism requires.

Acknowledgments

I am grateful to Kelly James Clark for helpful comments on a previous version of this paper.

References

Austin, J.L. (1950). Truth. *Proceedings of the Aristotelian Society* 24(Suppl.): 111–129. Reprinted in Austin, J.L. (1961). *Philosophical Papers*, edited by J.O. Urmson and G.J. Warnock. Oxford: Clarendon Press.

Block, N. and Stalnaker, R. (1999). Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review* 108: 1–46.

Brogaard, B. (Forthcoming). The Self-Locating Property Theory of Color. *Minds and Machines*.

Brogaard, B. and Salerno, J. (2005). Anti-Realism, Theism, and the Conditional Fallacy. *Nous* 39: 123–139.

Chalmers, D.J. (1996). *The Conscious Mind: In Search of a Theory of Conscious Experience*. New York: Oxford University Press.

Chirimuuta, M. (2015). Perceptual Pragmatism. In *Outside Color*. Cambridge, MA: MIT Press.

Cohen, J. (2009). *The Red and the Real*. Oxford: Oxford University Press.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Devitt, M. (1994). The Methodology of Naturalistic Semantics. *Journal of Philosophy* 91(10): 545–572.

Dowell, J.L. (2004). From Metaphysical to Substantive Naturalism: A Case Study. *Synthese* 138(2): 149–173.

Dummett, M. (1959). Truth. *Proceedings of the Aristotelian Society* 59: 141–162.

Field, H. (1972). Tarski's Theory of Truth. *Journal of Philosophy* 69: 344–375.

Field, H. (1994). Deflationist Views of Meaning and Content. *Mind* 103(411): 249–284.

Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Hill, C.S. (1997). Imaginability, Conceivability, Possibility, and the Mind–Body Problem. *Philosophical Studies* 87: 61–85.

Horwich, P. (1990). *Truth*. Oxford: Basil Blackwell.

Jackson, F., Oppy, G., and Smith, M. (1994). Minimalism and Truth Aptness. *Mind* 103(411): 287–302.

Koch, C. (2012). *Consciousness: Confessions of a Romantic Reductionist*. Cambridge, MA: MIT Press.

Kripke, S.A. (1980) *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly* 64: 354–361.

Lewis, D. (1970). How to Define Theoretical Terms. *Journal of Philosophy* 67: 427–446.

Loewer, B.M. (1997). A Guide to Naturalizing Semantics. In *A Companion to the Philosophy of Language*, edited by C. Wright and Bob Hale. Oxford: Blackwell.

Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Lynch, M. (2000). Alethic Pluralism and the Functionalist Theory of Truth. *Acta Analytica* 15: 195–214.

Lynch, M.P. (2009). *Truth as One and Many*. Oxford: Clarendon Press.
Moore, G.E. (1901/02). Truth and Falsity. Reprinted in Moore, G.E. (1993). *Selected Writings*, edited by T. Baldwin, pp. 20–22. London and New York: Routledge.
Papineau, D. (2009). Naturalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/spr2009/entries/naturalism/ (last accessed July 15, 2015).
Peirce, C.S. (1992). *The Essential Peirce*, vol. 1. Bloomington: Indiana University Press.
Plantinga, A. (1982) How to Be an Anti-Realist. *Proceedings and Addresses of the American Philosophical Association*, 56(1): 47–70.
Putnam, H. (1983). *Realism and Reason*. Cambridge: Cambridge University Press.
Quine, W.V.O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Quine, W.V.O. (1974). *The Roots of Reference*. La Salle: Open Court.
Ramsey, F. (1954). *The Foundations of Mathematics*. London: Routledge.
Rea, M. (2000). Theism and Epistemic Truth-Equivalences. *Nous* 34: 291–301.
Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
Rorty, R. (1981). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
Rorty, R. (1982). *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.
Rorty, R. (1985). Solidarity of Objectivity? In *Post-Analytic Philosophy*, edited by J. Rajchman and C. West. New York: Columbia University Press.
Rorty, R. (1991). *Philosophical Papers,* vol. 1: *Objectivity, Relativism and Truth*. Cambridge: Cambridge University Press.
Rorty, R. (1998). *Philosophical Papers,* vol. 3: *Truth and Progress*. Cambridge: Cambridge University Press.
Russell, B. (1904). Meinong's Theory of Complexes and Assumptions I, II, III. *Mind* 13: 204–219, 336–354, 509–524.
Russell, B. (1918/1956). The Philosophy of Logical Atomism. In *Logic and Knowledge: Essays 1901–1950*, edited by R. C. Marsh, pp. 177–281. London: George Allen and Unwin.
Sartwell, C. (2007). The Provocateur's Philosopher. *Los Angeles Times,* June 12, 2007. Available from: http://articles.latimes.com/2007/jun/12/opinion/oe-sartwell12 (last accessed July 15, 2015).
Siegel, S. (2010). The Contents of Perception. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/win2010/entries/perception-contents/ (last accessed July 15, 2015).
Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
Stoljar, D. and Damnjanovic, N. (2010). In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/win2010/entries/truth-deflationary/ (last accessed July 15, 2015).
Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
Wittgenstein, L. (1921/1961). *Tractatus Logico-Philosophicus*. London: Routledge.
Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
Wright, C. (2000). Truth as Sort of Epistemic: Putnam's Peregrinations. *Journal of Philosophy* 97(6): 335–364.
Zemach, E.M. (1987). Truth and Some Relativists. *Grazer Philosophische Studien* 29: 1–11.

20

Mathematics and Metaphysical Naturalism

GIDEON ROSEN

If naturalism is the vague idea that everything is part of nature – the world of matter and energy spread out in space and time – then there exists what appears to be an easy refutation of naturalism. There are, after all, infinitely many prime numbers. Euclid proved this, and the proof is as solid as can be. Moreover, it is obvious on reflection that prime numbers and the other objects of pure mathematics are not literally parts of nature. It is just plain silly to ask where the number 17 or the cosine function is, or how much they weigh, or how fast they're moving. So prime numbers and trigonometric functions exist – as mathematics shows – but they are not parts of nature – as is obvious. So naturalism stands refuted by elementary reflection on mathematics.

There are exactly three ways to resist this line of thought. One can deny that numbers and functions exist. One can allow that such things exist but insist that they are somehow parts of nature. Or one can refine the characterization of naturalism so as to allow numbers and other abstract objects as harmless exceptions that do not threaten what is important in the naturalistic view of the world.

The first approach has been discussed at length (Quine and Goodman 1947; Putnam 1970; Field 1980; Burgess and Rosen 1997). It amounts to denying the settled existence claims of mathematics, not because they are bad by mathematical standards, but rather because they offend some philosophical scruple. The philosopher who takes this approach will of course allow that Euclid's theorem is *acceptable*. It is not junk. We are entitled to speak and act *as if* it were true in every scientific and mathematical context. The distinctive claim of this approach is that a fully acceptable mathematical claim need not be true, and that the acceptance of a mathematical claim for serious purposes need not involve commitment to its truth.

While there is much to be said about this gambit, I propose to set it to one side. The key point against it comes from a position that has also (somewhat confusingly) been called "naturalism": the epistemological view that scientific claims are to be assessed by scientific standards, and not by some alien philosophical standard that has no place in scientific

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

practice (Maddy 1997). ("Naturalism" in this sense has nothing to do with *nature*. The word is appropriate only as a kind of pun: the philosopher is to be reckoned a *naturalized citizen* of the scientific community, someone who has made their standards her standards.) I endorse epistemological naturalism so conceived. In particular, I hold that when a mathematical claim is fully acceptable by mathematical standards (and by the standards of every other science), it simply follows that we are justified in believing that it is true (Rosen and Burgess 2007). A philosopher who favors an epistemic standard according to which the textbook proof of Euclid's theorem does not provide good reasons for believing Euclid's theorem is simply mistaken about what counts as a good reason for believing a mathematical proposition. He is like the skeptic who holds that the usual reasons for believing in atoms and molecules aren't good enough even though they clearly are good enough by the standards of chemistry and physics. I shall therefore assume in what follows that we cannot reasonably deny the existence theorems of standard mathematics. These claims are fully justified by mathematical standards, so they are justified *tout court*.[1] But that is just to say that we cannot deny the existence of numbers and other mathematical objects. The open question is what these things are like.

The second approach has not had many defenders in recent years. No one has ever seriously maintained that the objects of mathematics are ordinary physical objects like rocks or stars. One sometimes hears it said that numbers are *ideas in the mind*, and I suppose ideas might count as parts of nature, located in the minds or brains of concrete thinkers. But this view has never been developed as a serious philosophy of mathematics, for obvious reasons. In whose mind does the number 17 exist? Are there many 17s, one for each mind? If so, what are we to make of the theorem according to which there is exactly one natural number between 16 and 18? The only version of this view worth discussing would identify mathematical objects not with concrete ideas, but with the intended *objects* of those ideas: the thing we are thinking *of* when we entertain an idea of 17. But whatever its merits, this view cannot help the literal-minded naturalist, since even if the ideas themselves are parts of nature, it does not follow that the objects of those ideas are parts of nature. (You can't defend naturalism in theology by arguing that our *ideas* of God exist in nature.)

It has occasionally been maintained that mathematical objects are somehow *aspects* or *features* of natural things: universals that inhere in concrete particulars. On the best-developed versions of this view (Forrest and Armstrong 1987; Bigelow 1988), the cardinal numbers are identified with *relations* between concrete objects and properties. The number 3, for example, might be the relation that a heap of three apples bears to the property of being an apple (which is also the relation that a collection of three bears bears to the property of being a bear). Of course, if this is to be a naturalistic view, then all of these universals must be parts of nature, in which case we face a question about where and when they exist. Even if we grant that the property of being an apple exists *in* the apples that instantiate it – wholly present in this apple, and again in that apple, though not in any proper part of either – what shall we say about the *relation*, which this view identifies with the number 3, that the collection of apples bears to this property? Where is *it*? We can begin to invent answers, but at this point it should be clear that our metaphysical imagination has run away with us.

My own view, then, is that there is a flat contradiction between mathematics and metaphysical naturalism, so long as the latter is understood in the most straightforward way, viz., as the view that absolutely everything is located in the natural world of space and time.

1 See Rosen (2001) for a more nuanced statement of this view.

So one of the two must go, and the choice is clear. Standard mathematics is as well established as can be. Metaphysical naturalism is a philosophical picture. And while it is hard to make a firm rule that philosophy has no authority to overrule the sciences, its authority is clearly limited. If a philosopher wants to reject a well-established scientific claim on philosophical grounds, she needs a very good argument (Lewis 1991, 55ff.). But there are no good arguments against the existence of prime numbers and the cosine function. Simple-minded naturalism must therefore exit gracefully.

This brings us to the third strategy for reconciling metaphysical naturalism with mathematics. Perhaps the right moral to draw from this discussion is that the best version of naturalism is not the simple-minded view that requires everything to be located in space and time, but something weaker that preserves the spirit of this idea. What might this be? There are many possibilities, of which I will mention two.

Metaphysical naturalism derives its energy from its opposition to *supernaturalism*, the view that there are *causal agencies* that are not part of nature, but which impinge on nature: gods, transcendent forms of which the natural world is an emanation, a noumenal reality of which the natural world is a mere appearance, and so on. If we define naturalism (still vaguely) as the rejection of supernatural *causes*, then there is no tension between ordinary mathematics and naturalism. The cosine function, after all, is not a cause. It does not literally make things happen. If naturalism is understood as the rejection of supernatural causes, then the tension between mathematics and naturalism disappears.

Since the word "naturalism" has no established meaning, we are free to refine it in this way. But the choice has some surprising implications: Consider the view that the subject matter of mathematics is a world of invisible, immaterial, irreducible, eternal that never interact with one another or with the material world in any way. That would be consistent with naturalism on the present account, but it is precisely the sort of view from which most self-styled naturalists recoil.

Fortunately, there is an alternative refinement that is also worthy of our attention and which takes us in another direction. On this view, naturalism is not a thesis about *everything* that exists, but rather only about what exists *at the fundamental level*.[2] Bob and Phil are married, so their marriage exists. Their marriage is thus a thing: an item, an entity. Is that thing a part of nature? Is it located in spacetime? What an odd question! And who cares? To concede that Bob and Phil's marriage exists but that it doesn't really have a location is not to concede that the theists and the neoplatonists and the transcendental idealists were right in rejecting naturalism. This marriage isn't the sort of thing that needs to exist in nature in order for naturalism to be true. Why? Because if it exists at all, it is a *derivative* thing. The marriage exists *in virtue of* the fact that Bob and Phil exist and stand in certain relations to one another and to their surroundings. And since Bob and Phil and their surroundings are perfectly natural, and since they can stand in these relations simply by virtue of being suitably situated in nature, their marriage counts as natural even if *it* is not literally part of nature.

Let me try to make this more precise. The key idea is that some facts are *grounded in* – or *obtain in virtue of* – other facts.[3] The fact that Bob and Phil are married is grounded in a massive complex of social facts: facts about who said what to whom, and in what circumstances. The relation between the fact that Bob and Phil are married and this complex of

2 See Rosen (2011) for a somewhat fuller exposition of this idea.

3 See Rosen (2010) and the essays collected in Correia and Schnieder (2012) for extensive discussion.

social facts is very intimate. The fact Bob and Phil are married is not a *separate, free-floating fact*, though it is a *different* fact from the complex of facts that grounds it. One mark of this intimacy is this: If we could enumerate the massive complex of social facts in virtue of which Bob and Phil are married, we would see that this complex of facts *entails* that Bob and Phil are married. Given that the underlying social facts obtain, it follows necessarily that Bob and Phil are married. And in general: whenever *Q* grounds *P*, it follows that *Q entails P*.

In many cases in which *Q* grounds *P*, *Q* will entail *P but not vice versa*. The particular complex of social facts in virtue of which Bob and Phil are married guarantees that Bob and Phil are married. But the fact that Bob and Phil are married does *not* guarantee this particular complex of social facts. Bob and Phil might have been married in a different way – by a different rabbi, for example. There are, however, cases of grounding in which the entailment goes both ways. In these cases, *P* and *Q* are *equivalent* – each entails the other – and yet *P* is grounded in *Q*, but not vice versa. Consider a water molecule, *M*, made of three atoms, *H*, *H**, and *O*. The existence of the molecule is grounded in the fact that these three atoms exist and are bonded together in a certain way. And this means that the fact that the atoms are bonded in this way *entails* that the molecule exists. Conversely, the existence of *this* molecule, *M*, entails the existence of these atoms suitably arranged. (If these atoms had not existed, or had not been so arranged, this particular water molecule would not have existed.) So we have two propositions, each of which entails the other:

M exists,

and:

H, *H**, and *O* are arranged in such and such a way.

And yet it would clearly be a mistake to say that the atoms exist in this arrangement *because* the molecule exists. The only true explanatory claim in the vicinity runs the other way: the molecule exists because the atoms are suitably arranged. And that means that the existence of the molecule is grounded in the arrangement of the atoms and not vice versa.

In cases of this sort, we may say that the grounded fact – *M* exists – *reduces to* the grounding fact. The grounded fact and the fact to which it reduces are distinct, since one explains the other. And yet they are as intimately related as two facts can be.

It is an open question whether this grounding relation can be defined in more basic terms. But definable or not, it appears to be intelligible. The relation can be explained in the only way in which a basic relation can be explained: by giving examples and stating principles in which it figures. We have already seen some of the main principles: if *Q* grounds *P* then *Q* entails *P*; if *Q* grounds *P* then *Q* explains *P*. As for examples: When an object is scarlet (and therefore red), it is red in virtue of being scarlet. Fred is an uncle in virtue of the fact that his sister has a child. 17 is a prime number in virtue of the fact that its only factors are 1 and 17, and so on. (For an extended defense of the grounding idiom, see Rosen 2010.)

If this talk of grounding makes sense, then we can say this: There are two ways for an object to be *natural*. It can be a part of the spatiotemporal, natural world. Paradigmatic natural objects like rocks and stars and elementary particles are natural in this sense. But it can also exist wholly in virtue of the existence and arrangement of things that are natural in the first sense. Bob and Phil are (presumably) natural in the first sense, but their marriage

isn't. And yet their marriage *is* natural in the second sense. The existence of the marriage is wholly grounded in the arrangement of things in space and time. By contrast, the objects to which the naturalist is professionally allergic – God, transcendent forms, and so on – are natural in neither sense. They do not exist in ordinary space, nor do they exist wholly in virtue of the arrangement of spatial items. Rather, they exist nowhere in space, while nonetheless existing *in their own right*.

Since the existence of Bob and Phil's marriage is obviously compatible with metaphysical naturalism as it is normally understood, we should not identify naturalism with the thesis that everything exists in space. Rather, we should take naturalism to be the thesis that absolutely everything exists in space *or exists in virtue of the arrangement of things in space*.

We can give an equivalent formulation if we assume that there exists a "lowest level": a collection of *fundamental* things whose existence and arrangement serve to ground the existence of everything else. Now, it is unclear whether there is a fundamental level (Schaffer 2003). For all we know, there are infinite descending chains in which the existence of *F*s is grounded in the existence and arrangement of *G*s, which is in turn grounded in the existence and arrangement of *H*s, ad infinitum. But if we assume a fundamental level, we can characterize naturalism as the view that every *fundamental* thing is part of nature.

On this account, the hard question in the metaphysics of mathematics is not "Do numbers exist?" – of course they do – or "Do numbers exist in space?" – of course they don't. It is rather, "Are numbers (and other pure mathematical objects) fundamental entities?" Only if the answer is "yes" is there a clash between mathematics and metaphysical naturalism. For, if the answer is "no," then mathematical objects are derivative objects (like Bob and Phil's marriage); and derivative objects don't have to exist in space in order for naturalism to be true.

The question is interesting quite apart from its connection to naturalism. It invites us to explore the connection between the relation I have called "grounding" – the "in virtue of" relation – and the various forms of "construction" and "dependence" we encounter in the foundations of mathematics. It is often said, for example, that the real numbers are not fundamental entities, but are rather "constructed" from the rational numbers, which are in turn constructed from the integers. The simplest way to understand this proposal is as a set-theoretic identification, according to which each real number r is identified with a certain set of rational numbers – say, the set of rational numbers less than r – and each rational number with a set of integers. (For example, we might identify the rational number ¾ with the set {3, {4}}.)

Does a view of this sort entail that real numbers are nonfundamental? That depends on whether (non-empty) sets are fundamental. The view that they are not is strongly encouraged by the metaphors of stepwise construction that pervade set theory and are enshrined in the so-called "iterative conception" of set (Boolos 1971). We say that the set $\{a, b\}$ is *constructed* from or *based upon* or *generated from* its elements, a and b. But, of course, no one thinks that sets are literally built up through a temporal process that involves rearranging preexisting objects – as if sets were artifacts we somehow conjured into existence. So what exactly is the cash value of the metaphor? Just this, one might think: The set $\{a, b\}$ exists in virtue of the existence of the objects a and b.[4] To say that a mathematical object is a construction is just to say that it exists in virtue of the existence of other, more fundamental things.

4 More exactly, we might wish to say that the set $\{a, b\}$ exists in virtue of the existence of a and b, together with a *principle* of pure set theory, such as the principle that whenever x and y exist, their pair set $\{x, y\}$ exists.

Let us suppose for a moment that sets are indeed constructions in this sense, and that sets are therefore nonfundamental (they exist, but only in virtue of the existence of other things). It will then follow that the existence of the vast cumulative hierarchy of sets is no threat whatsoever to naturalism, even if the sets do not literally exist in nature. For naturalism only requires that *fundamental* things exist in nature, and sets, on this view, are not fundamental things.

One must instantly note a caveat concerning the empty set, $\varnothing$ – which is not "constructed" or "built up" from other sets, but rather serves as an unconstructed building block from which other sets are constructed. Standard mathematics assumes the empty set, so if it must be reckoned fundamental, and if *it* is not part of nature, then there would still be a clash – at this one isolated point – between set theory and metaphysical naturalism as we have defined it.

Now, when some important issue in the philosophy of mathematics seems to turn on the metaphysical status of the empty set, that's a good indication that our discussion has run off the rails. The empty set is a convenience, but it is thoroughly and trivially dispensable for all scientific and mathematical purposes (Burgess and Rosen 1997, 151). So if it were the only thing standing between us and a tenable naturalism, we could simply deny its existence and be none the worse for it.[5]

What about mathematical objects that are not sets? It was once conventional to identify all mathematical objects with sets, but, since Benacerraf (1965), it has been clear that when it comes to the natural numbers and other mathematical objects that are not given explicitly as sets, no determinate identifications of this sort are credible. (Which set is the number 17? There is no good answer to this question.) We should therefore consider the implications for metaphysical naturalism of the view that mathematics trades in nonspatiotemporal items that cannot be identified with sets.

Suppose, then, that the natural numbers 0, 1, 2,… are *sui generis*. If these numbers are fundamental, then metaphysical naturalism is false, since whatever these numbers are, they are obviously not parts of nature. So the question is whether numbers of this sort must be fundamental entities. If they are not, then every fact about these numbers – every *arithmetical* fact – must be grounded in facts that are not about numbers. The challenge for the mathematical naturalist is to make this plausible, ideally by explicitly identifying the nonarithmetical facts in virtue of which the arithmetical facts obtain.

Let's focus on *pure* arithmetical facts, facts like:

$2 + 3 = 5$.
For any two numbers n and m, $n + m = m + n$.
There are infinitely many prime numbers.
For all numbers a, b, and c, there is no $n > 2$ such that $a^n + b^n = c^n$.

The pure facts are just about the numbers, by contrast with *mixed* arithmetical facts, which are about relations between numbers and other things. (A complete defense of mathematical naturalism should cover both cases, but in the interest of time and space, we focus on the pure case.) The challenge is to describe a mapping from the facts of pure arithmetic to the

5 Alternatively, we could follow David Lewis (1991, 10ff.), who identifies the empty set (in effect) with the totality of. (To be more exact, Lewis identifies the empty set with the aggregate of all *individuals* – where an individual is anything that does not contain a set as a part.) This is obviously an arbitrary choice, but nothing in mathematics rules it out. And if we make it, the clash between set theoretic mathematics and metaphysical naturalism disappears (provided the individuals from which the null set is constructed are natural).

non-arithmetical facts that ground them. In order to see how this might be done, we need three well-known facts about arithmetic.

First fact: While mathematicians use an extensive technical vocabulary for convenience ("127," ">," "prime number"), almost all of this verbiage is dispensable in principle. If we have a name for zero ("0") and a symbol for the number that comes immediately after *n* (the successor of *n*, written "s(*n*)"), then we can dispense with the individual numerals "1," "2," "3," in favor of "*s*0," "*ss*0," "*sss*0," and so on. Instead of saying "5 > 3," we can say, "There is a number *x* such that 3 + *x* = 5," or, more longwindedly, "There is a number *x* such that *sss*0 + *x* = *sssss*0." Much less obviously, given suitable logical background, we can dispense not only with readily defined symbols like "3" and ">," but also with apparently basic mathematical symbols like "+" and "×."[6] At the end of the day, three arithmetical words suffice: "0," "number," and "successor." Given this minimal vocabulary, we can say everything that can be said about the natural numbers in the rich and sprawling language of number theory.

Second fact: Pure arithmetic can be axiomatized. Every true statement in the language of arithmetic is a logical consequence[7] of five basic statements, the so-called Dedekind–Peano axioms:

0 is a number.
0 is not the successor of a number.
Every number has a successor.
No two numbers have the same successor.
For any property *F*, if 0 has *F* and the successor of *n* has F whenever *n* does, then every number has *F*.

Third fact: These axioms can be used to define a *structure*, sometimes called an *omega sequence*. An omega sequence can contain anything whatsoever. It can be made of numbers:

0, 1, 2, 3, 4,…

or of people:

Adam, Cain, Enoch, Erad…

or of miscellaneous items. What matters is not the content of the sequence but its form. Every omega sequence consists of a bunch of objects together with a *relation*. In the case of the natural numbers, we have the relation "*y* is the successor of *x*." In the case of Adam and his descendants, the relevant relation might be "*y* is the first-born child of *x*."[8] What all omega sequences have in common is that relative to this ordering, the sequence is a *discrete linear ordering with a first element and no last element, in which any two elements are separated by a*

6 For the cognoscenti, I assume full second-order logic. In this framework, function symbols like "+" are explicitly definable in terms of "0," "number," and "successor."

7 This is not to say that every true statement can be proved from the axioms. As Gödel showed, if the axioms are consistent, then given any set of formal rules for deriving consequences from the axioms, there will always be true statements that cannot be derived. This is compatible with the claim in the text. The key point is that a statement can be a (semantic) logical consequence of the second-order axioms even if there is no formal proof of that statement from the axioms.

8 Adam and his first-born descendants form an omega-sequence only if the sequence never ends, since an omega-sequence must be infinite. So the example is good only under highly optimistic assumptions.

finite number of links. The Dedekind–Peano axioms define this structure in the following sense. Take these axioms and replace the distinctively mathematical words with letters:

x is an X.
x is not the R of an X.
Every X has an R.
No two Xs have the same R.
For any property F, if x has F and whenever an X has F, so does its R, then every X has F.

We can then say that whenever we have an object x, a collection X, and a relation R satisfying these conditions, the result is an omega sequence.

With all of this in place, we can entertain a natural hypothesis about the non-arithmetical facts that underlie the arithmetical facts – a version of what is sometimes called *structuralism* in the philosophy of mathematics.[9] Let A be any true claim about the numbers: "2 + 3 = 5," Fermat's Last Theorem,[10] whatever you like. To generate a plausible candidate for the nonarithmetical fact in virtue of which A obtains, we follow a recipe:

Step 1: Rewrite A in the stripped-down vocabulary in which the only mathematical words are "0," "number," and "successor."

The result – call it "$A[0, N, s]$" – may be long and hard to understand. But it says exactly what A says, and so expresses the same fact.[11]

Step 2: Let $PA[0, N, s]$ be a single sentence that conjoins the Dedekind–Peano axioms listed before. Then form the *conditional claim*:

If $PA[0, N, s]$, then $A[0, N, s]$.

This says that *if* the axioms are true, then our original statement is also true. Because the claim is hedged in this way, it does not entail the existence of numbers. It's the kind of thing a cautious philosopher might say: "I don't know whether there are infinitely many primes because I'm not sure numbers exist at all; but I'm happy to say that *if there are numbers satisfying the standard axioms, then there must be infinitely many primes*." Since the standard axioms logically imply every truth about the numbers, this conditional claim will be true whenever A is.

Step 3: Replace the arithmetical words in our conditional statement with letters and form a general claim as follows:

For any object x, collection X, and relation R: If $PA[x, X, R]$, then $A[x, X, R]$.

9 Arithmetical structuralism goes back to Dedekind (1887/1963). Benacerraf (1965) revived interest in the approach. The most important recent treatments are Hellman (1989) and Shapiro (1997). The current approach is closest to that of Kalderon (1996).

10 The famous claim, conjectured by Fermat in 1637 and finally proved by Andrew Wiles in 1994, that the equation $a^n + b^n = c^n$ has no solutions in the positive integers for n greater than 2.

11 This claim is not obvious. One might wish to distinguish the fact that 2 is the successor of 1 from the fact that $ss0 = ss0$ on the ground that the first fact includes the number 2 and the number 1 as *constituents* (along with the successor function), whereas the second contains only the number 0 and the successor function.

The resulting claim is not about numbers. It is a completely general claim, which says in effect: In any omega sequence – no matter what it's made of – a certain structural condition derived from the mathematical statement *A* holds. If we start with the claim "2 + 3 = 5," the result of the process is a claim to the effect that in any omega sequence, if you start with the third element[12] in the series and count up three, you reach the sixth element. If we start with the claim that there are infinitely many prime numbers, the result is a claim to the effect that there are infinitely many "prime elements" in any omega sequence. And so on. If we assume that there *are* omega sequences,[13] then these general conditional claims will be true exactly when the original mathematical claim *A* is true: each will entail the other. Moreover, the general claim preserves most of the mathematical content of the original. Whenever a mathematician proves a theorem about the natural numbers, she in effect proves a general theorem about omega sequences; and for most mathematical purposes, the original arithmetical claim and the general claim associated with it will be interchangeable. From a philosophical point of view, however, they are quite different. After all, one is a claim *about numbers* and the other isn't.

We are finally in a position to state a proposal that, if true, would reconcile arithmetic taken at face value (as a theory of *sui generis* nonspatial numbers) with metaphysical naturalism. The proposal is that while the claims of ordinary arithmetic are claims about nonspatial entities, these claims are *made true*, at a more fundamental level, by facts that are not about such objects: general conditionals that say, "If you've got an omega sequence, then it satisfies a certain structural condition." The metaphysical naturalist should be happy to acknowledge facts of this latter sort at the fundamental level, since they don't require the existence of supernatural entities. (In fact, they don't require the existence of anything: they are truths of logic.) So, if the pure arithmetical facts are systematically grounded in these innocuous facts, the existence of numbers is consistent with naturalism as we have construed it.

Let me close with two problems for this highly speculative approach. The strategy has been to associate each ordinary mathematical truth about numbers with an allegedly more fundamental truth consistent with naturalism. I've sketched one way of doing this. But, in fact, there are many ways of carrying out the strategy, each quite different from the others. Over the course of the 20th century, philosophers developed a variety of reductive proposals designed to demystify the objects of mathematics, each of which associates statements about numbers (and other mathematical objects) with statements in an allegedly more fundamental idiom that make no reference to numbers (and the rest). The strategies are all somewhat complicated, and it would take us too far afield to present them. Suffice it to say that each yields a mapping from statements about numbers to statements in a nonmathematical

12 Since we start with 0, 2 is the *third* element in the natural number series. Our procedure thus converts arithmetical claims about the number 2 into general claims about the third element of an arbitrary omega sequence.

13 This is a big if, or so one might think. An omega sequence is an infinite sequence, and we have no logical guarantee that there are infinitely many objects. For this reason, structuralist approaches to arithmetic often associate an arithmetical claim *A*, not with the generalized conditional claim we have been discussing, but rather with a stronger claim:

As a matter of *necessity*, for all objects *x*, collections *X*, and relations *R*, if PA[*x*, *X*, *R*], then *A*[*x*, *X*, *R*].

This claim is guaranteed to be equivalent to A on the assumption that there *could have been an omega sequence*. And unlike the claim that there actually *is* an omega sequence, that claim is certainly (and almost uncontroversially) true.

idiom, which maps true statements on to true statements, false statements on to false statements, and which appears to preserve much of the mathematical content of the original (see Burgess and Rosen 1997 for a survey.) So, for example, while the account sketched here maps the arithmetical claim "2 + 3 = 5" on to the conditional claim "In any omega sequence, if you start at the third element and count up three, you reach the sixth element," an alternative mapping would associate "2 + 3 = 5" with a rather different general claim:

If there are two *F*s and three *G*s and nothing is both *F* and *G*, then there are five things that are either *F* or *G*.

This is a mapping of the right kind: it associates a statement about numbers with a statement that does not require the existence of numbers for its truth. Extending this particular mapping to more complicated arithmetical claims is not straightforward, but it can be done (Hodes 1984). And the same trick can be pulled in many other ways.

Historically, these accounts were proposed as accounts of the *meanings* of mathematical statements, and for this purpose they are rarely plausible (Burgess 1983). They were also often intended to show that mathematical discourse is not really concerned with mathematical objects after all, and in this respect they were certainly mistaken: if number theory says that there are infinitely many prime numbers, then regardless of whether this claim can be "mapped" to a claim that doesn't mention numbers, number theory is committed, by this very claim, to the existence of infinitely many primes (Alston 1958). These proposals can, however, be reconceived as accounts of the facts that *ground* the mathematical facts, and for that purpose they may be fully adequate. And this raises a question: Given two accounts of the natural facts that allegedly ground a given mathematical fact, how are we to decide which is correct?

A related question would arise even if there were only one tenable proposal of this sort. Consider a debate between a naturalist who says that arithmetical facts are grounded in general conditional facts about omega sequences and a non-naturalist who thinks that the mathematical facts are metaphysically fundamental. These two views agree about the mathematical facts and about the general, conditional facts. Indeed, they agree that given any pure arithmetical statement *A*, *A* is true if and only if the associated general conditional claim is true. They disagree only on the philosophical question: Are the arithmetical truths *grounded* in the general conditional facts we have identified? And the question is: How can we choose between these two theories? What sort of *argument* would show that the mathematical facts are (or are not) grounded in the nonmathematical facts in this way?

One partial response to the first problem is to note that, faced with several candidates for the facts that ground a given mathematical fact, we are not necessarily forced to choose. A single fact can have many grounds. The fact that Bill is an uncle might be grounded *both* in the fact that his sister has a son and in the (independent) fact that his brother has a daughter. Similarly, it is at least a theoretical possibility that the mathematical fact that 2 + 3 = 5 is grounded *both* in the fact that every omega sequence satisfies a certain structural condition and in some quite different fact, such as the fact that whenever there are two *F*s and three *G*s that aren't *F*s, there are five things that are either *F* or *G*. Both of these proposals cite grounds that are intuitively more fundamental than the mathematical fact they allegedly ground and which seem to *explain why* the mathematical fact obtains. It is therefore not out of the question that the mathematical facts are grounded in the nonmathematical facts in several ways.

This observation also gives us a partial response to the second problem, the problem of choosing between naturalism of this sort and non-naturalism. According to the naturalist, every fact about mathematical objects has an explanation that ultimately makes no reference to mathematical objects. Asked *why A* is true, one can say everything that an ordinary mathematician would say, explaining *A* by reference to more fundamental mathematical truths that ground it. But when we come to the axioms, which have no explanation within mathematics, we can say more: the axioms hold because they are grounded in more fundamental facts that do not involve numbers. According to the non-naturalist, by contrast, an arithmetical fact that cannot be explained within mathematics simply cannot be explained at all. Now, the latter view is not incoherent. But given a choice between a credible explanation and no explanation at all, it is often reasonable to opt for the explanation. The main advantage of mathematical naturalism of the sort we have been discussing is precisely that it holds out the possibility of explanations of the mathematical facts from below. None of this amounts to a proof of naturalism or an argument for any particular naturalistic proposal. But it does suggest that the case for naturalism does not require supernatural capacities for discerning the invisible grounding structure of reality. Philosophers of science have long emphasized that it is often reasonable to prefer a theory that exhibits greater explanatory coherence than its rivals: a theory in which fewer facts need to be taken as unexplained. A naturalist theory of mathematics of the sort I have been describing may turn out to be more coherent in this sense than its non-naturalist rivals, since it will offer grounding explanations for mathematical truths that cannot otherwise be explained. And if that is so, it is a reason for taking the naturalistic proposal seriously.

And then there is the matter of intrinsic plausibility. In principle, this should count for very little in metaphysics, since there is no reason why our sense of intrinsic plausibility should be reliable when it comes to arcane questions about the basic structure of reality. But for what it's worth: I find it easy to believe that numbers and the rest exist as derivative entities – that they exist, but only because certain more fundamental facts obtain. I find it much harder to believe that numbers exist brutely, in their own right, alongside the basic constituents of physical reality, whatever they may be.

To review: Mathematics describes a realm of objects that are not literally parts of nature. That is why mathematics appears to pose a threat to naturalism. But the most interesting form of metaphysical naturalism is *not* the claim that absolutely everything is part of nature: it is claim that absolutely everything *at the fundamental level* is part of nature. Equivalently, it is the claim that when something is *not* part of nature, the facts about that thing are always grounded in facts about the natural universe. When naturalism is understood in this way, there is no clash between naturalism and mathematics. The clash disappears if we allow for the possibility that mathematical objects are derivative entities – like Bob and Phil's marriage – the facts about which are always grounded in facts about other things.

References

Alston, W. (1958). Ontological Commitments. *Philosophical Studies* 9: 8–17.
Benacerraf, P. (1965). What Numbers Could Not Be. *Philosophical Review* 74: 47–73.
Bigelow, J. (1988). *The Reality of Numbers*. Oxford: Oxford University Press.
Boolos, G. (1971). The Iterative Conception of Set. *Journal of Philosophy* 68(8): 215–231.
Burgess, J. (1983). Why I am Not a Nominalist. *Notre Dame Journal of Formal Logic* 24(1): 93–105.

Burgess, J. and Rosen, G. (1997). *A Subject with No Object: Steps Toward Nominalistic Reconstrual in Mathematics*. Oxford: Oxford University Press.

Correia, F. and Schnieder, B. (eds.) (2012). *Metaphysical Grounding*. Cambridge: Cambridge University Press.

Dedekind, R. (1887/1963). The Nature and Meaning of Numbers. In *Essays on the Theory of Numbers*, translated by Wooster Beman. Mineola: Dover.

Field, H. (1980). *Science without Numbers*. Princeton: Princeton University Press.

Forrest, P. and Armstrong, D.M. (1987). The Nature of Number. *Philosophical Papers* 16(3): 165–186.

Hellman, G. (1989). *Mathematics without Numbers*. Oxford: Oxford University Press.

Hodes, H. (1984). Logicism and the Ontological Commitments of Arithmetic. *Journal of Philosophy* 81: 123–149.

Kalderon, M. (1996). What Numbers Could Be (and Hence Necessarily Are). *Philosophia Mathematica* 4(3): 238–255.

Lewis, D. (1991). *Parts of Classes*. Oxford: Basil Blackwell.

Maddy, P. (1997). *Naturalism in Mathematics*. Oxford: Oxford University Press.

Putnam, H. (1970). *Philosophy of Logic*. New York: Harper and Row.

Quine, W.V. and Goodman, N. (1947). Steps Toward a Constructive Nominalism. *Journal of Symbolic Logic* 12: 105–122.

Rosen, G. (2001). Nominalism, Naturalism, Epistemic Relativism. *Philosophical Perspectives* 15 (*Metaphysics*): 69–91.

Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In *Modality: Metaphysics, Logic and Epistemology*, edited by Bob Hale and Aviv Hoffman, pp. 109–135. Oxford: Oxford University Press.

Rosen, G. (2011). The Reality of Mathematical Objects. In *Meaning in Mathematics*, edited by John Polkinghorne, pp. 113–131. Oxford: Oxford University Press.

Rosen, G. and Burgess, J. (2007). Nominalism Reconsidered. In *The Oxford Handbook of Philosophy of Mathematics and Logic*, edited by Stewart Shapiro, pp. 515–535. Oxford: Oxford University Press.

Schaffer, J. (2003). Is There a Fundamental Level? *Noûs* 37(3): 498–517.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.

21

Naturalism and Mathematics: Some Problems

JEFFREY W. ROLAND

Introduction

I have been asked to write in opposition to naturalism in mathematics. Not because I'm generally hostile to naturalism; I'm not. I tend to think that some sort of naturalism is probably right when it comes to empirical knowledge. And not because I have a well-developed non-naturalistic account of mathematics; I don't. Such an account is a tall order. But I'm working on it. I have been asked to write in opposition to naturalism in mathematics because I think that there are serious problems with naturalizing mathematics, understood as a project with both epistemological and metaphysical components, and I've been vocal about it (Roland 2007; 2008; 2009; 2013b).

On the face of it, mathematical knowledge and mundane knowledge of the physical world appear to be radically different. The former at least looks as if it should be amenable to a fairly straightforward story grounded in sensory experience (for medium-sized physical objects like tables and chairs, at any rate). After all, in lots of cases we can point to the thing we're interested in and perceive its properties without much fuss.[1] For mathematical knowledge, the story doesn't appear to be the same at all. We can't point to numbers and we can't perceive their properties in the ways we can those of ordinary physical objects. So one might think that naturalizing mathematics is simply a nonstarter.

The problem with looking at things in this way is that, at least since Locke, we have increasingly (and quite rightly) put a premium on integrating our knowledge into a broadly scientific worldview, a worldview where entities and faculties at odds with what we observe in the natural world are not to be countenanced. Moreover, the importance of integrating a particular type of knowledge is proportional to the role it plays in explaining our experience of the natural world. Thus, for example, many scientists and scientifically minded

1 In fact, things are a lot more complicated than I'm making them out to be here. But I'm presently just interested to get the intuitive contrast between mathematical knowledge and mundane knowledge of the physical world on the table.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

philosophers have been more than happy to jettison religious knowledge involving souls and deities. The situation with mathematical knowledge, however, is quite different. Very few scientists or scientifically minded philosophers are happy to jettison mathematical knowledge. The need to somehow integrate mathematical knowledge into a broadly scientific worldview has been and remains intensely pressing. This is where naturalizing mathematics comes in.

Naturalism is one of those things that philosophers can't seem to help disagreeing about. Some philosophers think it's perfectly obvious what naturalism is, while others think that 'naturalism' is such a term of art that it's hopeless to try and determine what it means. My own view is that, though 'naturalism' means different things to different philosophers, there is a reason that "naturalistic" positions are so identified – a motivating spirit of naturalism, if you will – which gives us some purchase on what naturalism is.[2] That said, the issues are sufficiently complex that I won't try to say anything comprehensive in this chapter about naturalism or how mathematics might be naturalized. Rather, I'll first say something about how the project of naturalizing mathematics should be understood, given what motivates philosophical naturalism more generally. I'll then discuss two of the leading approaches to naturalizing mathematics and some problems with these approaches. I'll end by tying together some threads common to the problems I've highlighted.

Motivations for Naturalism

Naturalism has become something of a default position in Anglo-American philosophy (henceforth, simply 'philosophy') in the sense that, given a naturalistic view of a field of study and a non-naturalistic view of the same field, all else equal, naturalism wins out over non-naturalism. The impulse toward naturalism which has led to this state of affairs can be attributed to two factors, one positive and one negative. The positive factor is the striking success of the sciences,[3] in particular the natural sciences.

As a source of explanation of and advantageous interference in the world, modern science is without equal. This fact leads us to think that we might well attain similar success in fields of study traditionally held outside the purview of science using the methods of science. An obvious objection to this thought is that the relevant fields of study have traditionally been held outside the purview of science for a reason, viz., they involve entities[4] traditionally thought not to belong to the natural world, the rightful domain of science – entities such as mental states and moral norms. The naturalist answers that the relevant entities are in one way or another tied up with the natural world so as to make the methods of science applicable (e.g., mental states are states of persons, who are part of the natural world, and what makes an action morally permissible or obligatory is facts concerning persons and their relations to other parts of the natural world).

Here, the negative impetus toward naturalism – the rejection of supranatural entities and faculties – comes into play. Paradigm cases of supranatural entities are deities and souls of the Judeo-Christian-Islamic sort and Cartesian minds – incorporeal thinking substances.

2 For what I think is required for naturalism, at least in the tradition of Quine, see Roland (2013a).

3 The term 'science' and cognates in this chapter should be understood to include the natural and social sciences and to exclude mathematics.

4 I use 'entity' and cognates to cover both objects and properties.

A paradigm case of a supranatural faculty is mathematical intuition, a quasi-perceptual faculty giving us access to mathematicalia in the abstract realm of Plato's heaven. The motivation here is empiricist: what there is is what we are able to get on to either directly via sensory experience or indirectly owing to its relation to entities accessible directly via sensory experience. Supranatural entities and faculties meet neither of these criteria.

We can identify different aspects of naturalism in these twin motivations for naturalism, one methodological and one metaphysical. The methodological aspect comes out in the move to apply the methods of science to *prima facie* nonscientific fields of study. This aspect is often also thought of as epistemological, since the methods of science are used to acquire knowledge in the fields of study to which they are applied. The metaphysical aspect comes out in the rejection of supranatural entities. This aspect is often thought of as ontological, since it concerns what there is. Distinguishing these aspects of naturalism leads some to think in terms of different kinds, as opposed to different aspects, of naturalism – methodological/epistemological and metaphysical/ontological[5] – as follows. Where $\mathcal{F}$ is some field of study concerning *F*s and their properties, we have:

Epistemological Naturalism about $\mathcal{F}$ ($EN_{\mathcal{F}}$): The methods of $\mathcal{F}$-inquiry are answerable to science in that methods of inquiry in $\mathcal{F}$, at a minimum, must be consonant with the methods of science, where this includes being a legitimate source of truths concerning *F*s and their properties by the lights of our best science.

Metaphysical Naturalism about $\mathcal{F}$ ($MN_{\mathcal{F}}$): The entities at issue in $\mathcal{F}$ are entities of the sort countenanced by our best science.

Of course, our concern in this chapter will be with $\mathcal{F}$ = mathematics. Before turning to the mathematical case, however, let's consider why $EN_{\mathcal{F}}$ and $MN_{\mathcal{F}}$ should be thought of as aspects of a thoroughgoing naturalism instead of as distinct kinds of naturalism.

For a given field of study $\mathcal{F}$, it's logically possible to endorse $EN_{\mathcal{F}}$ and not to endorse $MN_{\mathcal{F}}$. That is, there is no *logical* problem with (no contradiction in) holding that the right way to find out about some supranatural entities is via methods consonant with those of science. Similarly, it's logically possible to endorse $MN_{\mathcal{F}}$ and not to endorse $EN_{\mathcal{F}}$. Hence, one might be an epistemological naturalist about $\mathcal{F}$ without being a metaphysical naturalist about $\mathcal{F}$, and vice versa. But though this is logically consistent, it's problematic.

The issue is that one can't comfortably endorse $EN_{\mathcal{F}}$ and not also endorse $MN_{\mathcal{F}}$, or vice versa. The notion of comfort operative here is admittedly vague, but intuitively the idea is that holding one view while rejecting the other should not be intellectually awkward, in that holding the one would give a rational, informed person a nontrivial reason to also hold the other.

Consider first endorsing $MN_{\mathcal{F}}$ without endorsing $EN_{\mathcal{F}}$. If one believes that *F*s and their properties are natural entities, entities of the kind countenanced by our best science, then there is something decidedly odd about believing that *F*s and their properties are not best learned about using methods consonant with the methods of science. Indeed, one might reasonably think that the methods of science themselves are the right methods to use. After all, the methods of science, whatever they happen to be, are at least *prima facie* the best way we have of learning about natural entities. So one can't comfortably endorse metaphysical naturalism about $\mathcal{F}$ without endorsing epistemological naturalism about $\mathcal{F}$. Now consider

5 Hereafter, I will use 'epistemological' and 'metaphysical'.

endorsing $EN_{\mathcal{F}}$ without endorsing $MN_{\mathcal{F}}$. Supposing one believes that the best way to find out about *F*s and their properties is by using methods consonant with the methods of science, a very good reason for holding this belief would be believing that *F*s and their properties are entities of the sort countenanced by science, since this would make them prime candidates for being epistemically accessible via methods similar to those of science. Of course, it's possible to reject this explanation for the belief that the best way to find out about *F*s and their properties is by using the methods consonant with those of science, but it's not at all obvious what better alternative explanation might be offered. So epistemological naturalism about $\mathcal{F}$ can't comfortably be endorsed without also endorsing metaphysical naturalism about $\mathcal{F}$.

Since neither $EN_{\mathcal{F}}$ nor $MN_{\mathcal{F}}$ can comfortably be endorsed without endorsing the other, $EN_{\mathcal{F}}$ and $MN_{\mathcal{F}}$ should be thought of as aspects of a thoroughgoing naturalism incorporating both of them, rather than as distinct kinds of naturalism which we can take or leave as we like.[6] Thus, in this chapter, I will take it that a satisfactory naturalism is thoroughgoing in that it incorporates both epistemological and metaphysical components. Specific to mathematics, we can characterize this thoroughgoing naturalism by:

Thoroughgoing Naturalism about Mathematics: The methods of mathematics are answerable to science and the entities at issue in mathematics are entities of the sort countenanced by our best science.

This is the notion of naturalism with which we'll be working in the remainder of this chapter.

In the next two sections,[7] I discuss and critique two leading approaches to naturalizing mathematics (i.e., to giving a naturalistic account of mathematics). The first approach is really a family of related approaches, each having in common with the others that it is based on some variety of indispensability argument for mathematical realism. The second approach is the one advanced by Penelope Maddy as part of her *Second Philosophy*.[8] These approaches don't exhaust the ways in which one might attempt to give a naturalistic account of mathematics. They are sufficiently different and influential, however, that considering them both should allay worries that I've chosen to consider the one problematic outlying approach to naturalizing mathematics. If both of these approaches are in trouble, then that raises worries about the project of naturalizing mathematics itself.

Indispensability-Based Mathematical Naturalism

Quine–Putnam Indispensability Argument

Probably the dominant approach to naturalizing mathematics for the past half century has been based on the well-known Quine–Putnam indispensability argument for mathematical realism.[9] According to that argument, anything that plays an indispensable role in a well-confirmed theory of empirical science (broadly construed to include social sciences as well

6 I'm certainly not the first or only person to argue that a satisfactory naturalism should incorporate both epistemological and metaphysical components. See, e.g., Kornblith (1994), Rosenberg (1996), and Stroud (1996).

7 Significant portions of which are drawn from Roland (2007; 2009).

8 See, e.g., Maddy (1997; 2005a; 2007; 2011).

9 For Quine's contribution see, e.g., (1963; 1981). For Putnam's, see (1971; 1975).

as natural sciences) enjoys the same confirmation as the theory itself. Since mathematics plays an indispensable role in our best science, itself well confirmed, mathematics is well confirmed. Our best science commits us to electrons and gravity (among other things) because electrons and gravity are essential to our best science doing the work that it does and our best science is well confirmed. By the same token, mathematicalia such as numbers and functions are essential to our best science doing the work that it does. Thus, we're as justified in believing that there are numbers and functions as we are in believing that there are electrons and gravity.

There are a number of ways of cashing out the Quine–Putnam argument. Mark Colyvan (2001, 11), for instance, offers (and defends):[10]

(1) We are ontologically committed to all and only those entities that are indispensable to our best scientific theories.
(2) Mathematical entities are indispensable to our best scientific theories.
(3) Therefore, we are ontologically committed to mathematical entities.

Though the Quine–Putnam indispensability argument typically focuses on existential claims of mathematics (e.g., 'there are numbers'), these aren't the only statements of mathematics we're supposed to be committed to based on the indispensable role of mathemat ics in our best science. We're supposed to be similarly committed to any statement of mathematics that is indispensable to our best science. So it's by reference to our best science that we see which mathematical claims we're justified in believing and which mathematical entities we're justified in believing there are. This is mathematics answerable to science epistemologically and metaphysically, and this is how the Quine–Putnam indispensability argument appears to give us a way of naturalizing mathematics.

Despite its initial plausibility, the Quine–Putnam indispensability argument is subject to a fatal objection as a naturalistic account of mathematics. Seeing this requires introducing some constraints on a satisfactory naturalistic account of mathematics. These constraints will also play a role in evaluating the other versions of mathematical naturalism considered in this chapter.

I take it that to be a satisfactory view of a certain type requires respecting the motivations for that type of view. Thus, we can read (some) constraints on the satisfactoriness of a type of view off the motivations for views of that type. With respect to naturalism, this immediately gives us two constraints:

Epistemological Constraint (EC): An adequate naturalism must respect the epistemological motivation for naturalism, viz., the motivation to model methods of inquiry on the best available such methods, those being the methods of our best science.

Metaphysical Constraint (MC): An adequate naturalism must respect the metaphysical motivation for naturalism, viz., the motivation to abjure supranatural entities and faculties, those being entities abjured by our best science.

Another constraint comes from the general antiskeptical attitude of naturalism. Naturalism is a project of accounting for epistemologically successful enterprises: Given that we possess knowledge in some domain, how do we naturalistically explain our acquisition

10 I've made minor changes.

and maintenance of that knowledge? In light of the standard truth condition on knowledge (viz., one knows that p only if p is true), this antiskepticism leads to the:

Truth Constraint (TC): An adequate naturalism about $\mathcal{F}$ must (at least implicitly) ratify the truth of known propositions of $\mathcal{F}$.

I take implicitly ratifying the truth of known propositions of $\mathcal{F}$ minimally to be a matter of having the resources to tell a satisfying story about why the known propositions of $\mathcal{F}$ are true.

Finally, we have the:

Integrity Constraint (IC): An adequate naturalism about $\mathcal{F}$ must respect the integrity of $\mathcal{F}$, in the sense that it must not recommend revisions to $\mathcal{F}$ for reasons not of antecedent concern to $\mathcal{F}$.[11]

So, for instance, revising biological theory in response to a conflict with chemical theory would be naturalistically legitimate because avoiding conflict with our best chemistry is something biology already aims to do. On the other hand, revising biological theory in response to conflict with the Genesis story would not be naturalistically legitimate, because avoiding conflict with religious texts is not among the aims of biology.

Note that IC is, as Maddy suggests, fundamental to the spirit of naturalism:

> To judge mathematical methods from any vantage-point outside of mathematics, say from the vantage-point of physics, seems to me to run counter to the fundamental spirit that underlies all naturalism: the conviction that a successful enterprise, be it science or mathematics, should be understood and evaluated on its own terms, that such an enterprise should not be subject to criticism from, and does not stand in need of support from, some external, supposedly higher point of view. (1997, 184)

I've argued elsewhere[12] that, with respect to mathematics, Maddy takes this view too far, and I'll replay some of those arguments later in the chapter. For present purposes, what's important is that whether it's the idea that science need not look outside itself for validation of its methods and findings or the idea that mathematics need not look outside itself for validation of its methods and findings, the overarching idea is that an epistemically successful practice *prima facie* need not be responsive to the concerns of practices other than itself. Intrusion into the affairs of a practice from without is only legitimate where the concern driving the intrusion is one antecedently recognized by the practice as important to its aims. The upshot is that, in naturalizing mathematics, recommending revision to established mathematics merely to serve the naturalizing project is illegitimate in a way that compromises the naturalizing project itself.

Returning now to mathematical naturalism based on the Quine–Putnam indispensability argument, it's not hard to see that EC and MC aren't a problem. TC and IC, however, are a different story.

The problem is basically the same with respect to both. The amount of mathematics needed to do our best science turns out to be very small, relative to how much mathematics there is. So, on the Quine–Putnam indispensability approach, we have little justification for believing claims of mathematics that go beyond those which are indispensable to our best science and little reason to believe that the entities at issue in these claims exist. Quine was aware of this:

11 A defense of IC can be found in Roland (2009, § 1.2). There, (PI*) is essentially IC.

12 See Roland (2007).

> Pure mathematics, in my view, is firmly embedded as an integral part of our system of the world. Thus my view of pure mathematics is oriented strictly to application in empirical science. Parsons has remarked, against this attitude, that pure mathematics extravagantly exceeds the needs of application. It does indeed, but I see these excesses as a simplistic matter of rounding out…I recognize indenumerable infinities only because they are forced on me by the simplest systematizations of more welcome matters. Magnitudes in excess of such demands, e.g., $\beth_\omega$ or inaccessible numbers, I look upon only as mathematical recreation and without ontological rights. (1998, 400)

Not worrying ourselves with the details of what $\beth_\omega$ or inaccessible numbers are, the takeaway is that a lot of established mathematics gets left out on the Quine–Putnam indispensability approach. This violates TC because much of what we take to be mathematical knowledge fails to be true on this approach; it violates IC because that mathematics which now fails to be known should be excised, and this would be an illegitimate revision of mathematics.

Colyvan's Extended Indispensability Argument

It will be helpful to have a little terminology here. We'll call the mathematics justified or confirmed by the Quine–Putnam indispensability approach *indispensable mathematics*. The mathematics that gets left out by this approach we'll call *dispensable mathematics*.

One way of thinking about the problem with the Quine–Putnam indispensability approach is to consider how the approach succeeds (let's suppose) in naturalizing indispensable mathematics and how it fails to naturalize dispensable mathematics. The Quine–Putnam indispensability approach succeeds in confirming only those parts of mathematics required by empirical science. Naturalistic epistemology aims to account for our knowledge in terms of naturally explicable facts and faculties: facts and faculties studied by or otherwise compatible with the empirical sciences. The final arbiter of the empirical sciences is perceptual experience. So, naturalistic justification is a matter of there being a sufficiently strong connection to perceptual experience. The empirical sciences are directly confirmed because they directly confront, explain, and predict experience. Indispensable mathematics is confirmed indirectly by its (direct) role in the empirical sciences. Dispensable mathematics fails to be confirmed because it appears to be insufficiently strongly connected to experience.

This suggests that the shortcoming of the Quine–Putnam indispensability approach as a strategy for naturalizing mathematics as a whole might be overcome by arguing for a sufficiently strong connection between dispensable mathematics and experience. Colyvan (2001) gestures at such an argument:

> As for the charge that the [Quinean] indispensability argument leaves too much mathematics unaccounted for (i.e., any mathematics that does not find its way into empirical science), this seems to misrepresent the amount of mathematics that has directly or indirectly found its way into empirical science. On a holistic view of science, even the most abstract reach of mathematics is applicable to empirical science so long as it has applications in some further branch of mathematics, which may in turn have applications in some further branch until eventually one of these find applications in empirical science. Indeed, once put this way it's hard to imagine what part of mathematics could possibly be unapplied. (2001, 107)

The idea here is that dispensable mathematics is tied to experience via a chain of indispensable mathematical applications that culminates in an indispensable application to indispensable mathematics.

Call the sense in which a part of dispensable mathematics is indispensable to empirical science in virtue of its participation in a chain of indispensable applications culminating in an indispensable application to a part of indispensable mathematics *extended indispensability*, and call Colyvan's argument in the passage just quoted the *extended indispensability argument*. If the extended indispensability argument is correct, then it arguably fixes the problem with the Quine–Putnam approach to naturalizing mathematics. But the extended indispensability argument has its own problems. I focus on the most serious.

The most reasonable way of understanding one part of mathematics as having applications in another is in terms of the first being used to prove results in the second. But if we understand one part of mathematics as having applications in another part of mathematics proof-theoretically in this way, then work on predicativist mathematics by Solomon Feferman and others[13] shows that the extended indispensability argument is quite likely wrong.[14]

Developing the project started by Weyl (1918/1994), Feferman has shown that all mathematics that is indispensable to our best science can be proved using only arithmetic; that is, the formal theory of the natural numbers $(0, 1, 2, \ldots)$.[15] This implies that indispensable mathematics (in the sense of the Quine–Putnam argument) contains nothing beyond the natural numbers. Everything else is dispensable, and any appearance otherwise is mere appearance.[16] If Feferman is right, then the problems raised for the Quine–Putnam approach are even more serious than we thought. But more to the point here, since arithmetic gets us all we need for science, parts of mathematics other than arithmetic simply don't play the proof-theoretic role the extended indispensability argument requires of them. Even if Feferman turns out to be wrong, however, Colyvan's strategy faces other difficulties.

The core of the difficulty based on Feferman's predicativist mathematics just noted is that the amount of mathematics that gets used in empirical science is limited: that science doesn't need all of contemporary mathematics to do its work. There are other ways to argue for this limited-use claim.

Investigations in the philosophy and foundations of mathematics often turn to set theory, as ours now does. Intuitively, we can think of sets as collections or aggregates and of set theory as the study of sets. Some sets consist of concrete things (e.g., the set of planets in our solar system). Some sets consist of abstract things (e.g., the set of natural numbers). The set of planets in our solar system is finite, but (crucially) sets can be infinite. The set of natural numbers, for instance, is infinite. In fact, set theory is extremely rich in infinite sets: there are more different sizes of infinite sets than there are sets themselves.[17]

Set theory is a useful tool in these investigations because, since all of classical mathematics can be modeled there, working with set theory gives us a nice, clean way of talking about all of classical mathematics at once. Presently, we're interested in whether or not empirical science requires all of mathematics to do its work. So we consider this question in the form

13 See, e.g., Feferman and Jäger (1993; 1996), Ye (2000), and the papers in Part V of Feferman (1998). For a nice overview of the project, see Feferman (2005).

14 Colyvan (2007, 115) acknowledges that this work might be a problem for his extended indispensability view.

15 See Feferman (1988) for details.

16 Cf. Feferman's (1992) remarks concerning the ramifications of the predicativist program for the standard indispensability argument.

17 In the jargon of set theory: there are proper class many sizes of infinite sets.

of a question about set theory: Does empirical science require all of set theory to do its work? When I talk about set theory in this chapter, I have in mind the standard presentation of set theory, ZFC, without large cardinals.[18]

Now, here's the thing. It's a theorem of ZFC that, for any finite list of axioms of ZFC, there is a fragment of the set-theoretic universe in which every member of the list holds.[19] That is, intuitively, no matter what's on the list, we don't need anywhere near all sets to make everything on the list true. This fact limits the amount of set theory that is indispensable to empirical science in the extended sense.

To see this, let $S = \{\tau_1, \tau_2, \ldots, \tau_k\}$ be the set of mathematical theorems that are directly indispensable (i.e., indispensable apart from the extended sense) to empirical science. S is finite. At any given time, there are only finitely many statements of empirical science, and arguments for those statements use only finitely many mathematical results. Moreover, since each τ_i has a proof from the axioms of ZFC, there is a set $AX(S)$ of ZFC axioms from which every member of S is provable. But now, since proofs are finite and there are only finitely many τ_i, $AX(S)$ is finite and there is some fragment of the set-theoretic universe in which every member of $AX(S)$ holds, by the theorem noted at the beginning of the last paragraph. Thus, only entities in (or with set surrogates in) this fragment are required for empirical science, which implies that not (nearly) all of set theory is required for our best empirical science. Following the details here isn't important. What is important is that no matter how much set theory is needed to prove the results that are indispensable to empirical science in the extended sense, it will be only a small fraction of the whole of set theory.

Colyvan has recently been defending recreational mathematics, by which he means mathematics without (extended) indispensable application in our best science, and the naturalist might hope that the arguments used in this defense would be helpful in response here. Colyvan's basic idea is that "[m]athematical recreation is an important part of mathematical practice" and that "like other forms of theoretical investigation, [it] should not be thought of as second class or *mere* recreation" (2007, 116). We needn't worry over the details, because there are at least two reasons independent of them that the naturalist's hope for help here would be in vain.

First, Colyvan's recreational view is an extension of the view based on the extended indispensability argument we've been criticizing. So, the critique the naturalist would like help with applies equally well to the recreational view. Second, though Colyvan's extended indispensability argument confers ontological rights to much more mathematics than does the Quine–Putnam indispensability argument, the mathematics that escapes the extended indispensability argument, viz., recreational mathematics, is still without ontological rights. Colyvan "[accepts] that applied mathematics should be treated realistically and with unapplied [i.e., recreational] mathematics we have no reason to treat it this way" (2007, 116). Hence, recreational mathematics isn't truth apt and so is still excluded from the pool of mathematical knowledge. Thus, the same problem that led to Colyvan's extension of the indispensability argument arises anew: excluding significant parts of pure mathematics in this way violates TC and IC. So, the recreational view provides no help to the naturalist.

18 'ZFC' denotes Zermelo–Fraenkel set theory with the axiom of choice. For a textbook treatment of ZFC, see Enderton (1977) or Kunen (1980). For large cardinals, see Kanamori (1994).

19 This follows from the Reflection Theorem. See, e.g., Kunen (1980, ch. IV, § 7).

Maddy's Second Philosophical Naturalistic Mathematics

The most prominent advocate of naturalism in mathematics in recent years has been Maddy.[20] Though she now generally avoids labeling her view with the term 'naturalism', it is no less intended as a version of naturalism than it was when she called it 'mathematical naturalism'. These days, Maddy splits her view into *Second Methodology of Mathematics* (SMM) and *Second Philosophy of Mathematics* (SPM).

According to SMM, the methods of pure mathematics are the business of mathematics and mathematics alone.[21] Adjudicating the proper methods of pure mathematics – "What makes for a good definition, an acceptable axiom, a dependable method of proof?" (Maddy 2007, 344) – is purely internal to mathematics. In short:

> Actual methodological decisions…are based on a perfectly rational style of means–ends reasoning: the most effective methods available toward the concrete mathematical goals in play are the ones endorsed and adopted. (2007, 349)

So, for example, if we want to know whether to adopt a new axiom for set theory, we first identify specific mathematical goals of set theory and then ask whether adopting that axiom would further those goals more than adopting some other axiom candidate or adopting nothing new at all. If the answer is 'yes', then we're justified in adopting the axiom in question, and we should do so; otherwise, we're not justified in adopting the axiom in question, and we should refrain from doing so. The point is that decisions such as this one are made solely on mathematical grounds. Philosophical considerations have no place in these decisions:

> [T]raditionally philosophical questions of ontology and epistemology are irrelevant to the methodology of mathematics. (2007, 361)

But neither do scientific considerations:

> To judge mathematical methods from any vantage-point outside mathematics, say from the vantage-point of physics, seems to me to run counter to the fundamental spirit that underlies all naturalism: the conviction that a successful enterprise, be it science or mathematics, should be understood and judged on its own terms, that such an enterprise should not be subject to criticism from, and does not stand in need of support from, some external, supposedly higher point of view…Where Quine takes science to be independent of first philosophy, my naturalist takes mathematics to be independent of both first philosophy and natural science (including the naturalized philosophy that is continuous with science) – in short, from any external standard. (Maddy 1997, 184)

Mathematics and its methods are autonomous from philosophy, from the empirical sciences, from "any external standard." But the methods of mathematics aren't the whole story.

20 See the references in note 8.

21 I restrict my attention to Maddy's discussion of pure mathematics here, since pure mathematics is where issues concerning naturalism and mathematics tend to be the most pressing.

Maddy recognizes that, despite the irrelevance (by her lights) of traditional philosophical questions to adjudicating the proper methods of mathematics, those questions still must be faced. Mathematics, however, is ill-equipped to address them:

> Does mathematics have a subject matter, like physics, chemistry, or astronomy? Are mathematical claims true or false in the same sense? If so, by what means do we come to know these things? What makes our methods reliable indicators of truth? The answers to these questions will not come from mathematics itself (Maddy 2007, 361)

Thus does Maddy find room for *philosophy* of mathematics, though philosophy of mathematics subordinate to *methodology* of mathematics.

> [Answers to these questions] are constrained by the methodological autonomy of mathematics; they must sustain its methodological verdicts. (2007, 361)

Maddy offers two philosophies of mathematics –*Thin Realism* and *Arealism* – though she argues that, ultimately, one is as good as the other.

Thin Realism can be understood in terms of how it answers the questions. According to Thin Realism, set theory does have a subject matter, viz., sets;[22] set-theoretic claims have determinate truth values in virtue of facts about sets; we come to know this and other things about sets by deploying the methods of set theory; and these methods are reliable indicators of truth concerning sets because sets are exactly those things about which set-theoretic methods are reliable indicators of truth. For the Thin Realist, Maddy tells us, there is no "gap between the methods of set theory and sets: sets just are the sort of thing that can be known about in these ways" (2007, 370). Set theory tells us there are sets, so there are. Set theory endorses classical logic, and hence the law of excluded middle,[23] with respect to set-theoretic claims. So set-theoretic claims have determinate truth values. And so on. In short, on Thin Realism there is nothing more to be said about sets than what set theory itself says about sets.

In contrast, where Thin Realism is "predicated on the assumption that [a naturalistic philosopher] has good reason to regard pure mathematics in general, and set theory in particular, as a body of truths" (Maddy 2011, 88), Arealism rejects this assumption. Instead, Arealism holds that "[s]et theory is [merely] the activity of developing a theory of sets that will effectively serve a concrete and ever-evolving range of mathematical purposes" (2011, 89): set theory has no subject matter; set-theoretic claims are not truth apt and so have no determinate truth values; and, in the absence of truth aptness, knowledge concerning sets and the reliability of set-theoretic methods become moot.

Thin Realism and Arealism are both constrained by SMM to respect the methods of mathematics in matters of mathematical epistemology and metaphysics. Where they differ is in what they think the actual methods of mathematics imply about the metaphysics and epistemology of mathematics, specifically set theory. Thin Realism takes the methods of set theory to yield the existence of sets and knowledge about them; Arealism takes set theory

22 When Maddy gets into details concerning SPM, she tends to switch from considering mathematics generally to considering set theory.

23 For our purposes, we can think of the law of excluded middle as saying that every sentence is either true or false.

and its methods to be engaged in a different kind of enterprise, one in which truth and existence aren't relevant – so no sets. In either case, the entities of mathematics are just those countenanced by our best mathematics (set theory).

Maddy's contention that Thin Realism and Arealism actually come to the same thing notwithstanding, what we have here are two putative naturalistic accounts of mathematics, both versions of Maddy's Second Philosophical approach. One account weds SMM to Thin Realism; the other weds SMM to Arealism. Let's call these accounts SMM + TR and SMM + A, respectively. Neither SMM + TR nor SMM + A succeeds in naturalizing mathematics.

Consider first that neither account obviously satisfies EC. Whether or not EC is satisfied by either account depends on complicated issues to do with the relation between the methods of natural and social science and the methods of mathematics. But Maddy would likely object to her view being subject to EC on grounds that the methodological autonomy she accords mathematics makes it inappropriate to require the methods of mathematics to be answerable to those of science. Thus, Maddy might think there's some question begging in requiring her position to honor EC. I think that exempting her view from EC in this way undercuts her ability to tap into the considerable *prima facie* weight given to naturalistic positions by default, since that weight derives from the attractiveness of positions satisfying EC. But I suspect Maddy would disagree with me on this. So let's leave EC aside and consider how SMM + TR and SMM + A fare with respect to the other constraints. We begin with SMM + A.

There's no reason to think that SMM + A doesn't satisfy MC. Arealism doesn't obviously endorse any metaphysics at all, and given that mathematics isn't truth apt according to Arealism, there's no pressure to introduce metaphysics to help account for the truth values of mathematical claims. But just this lack of truth aptness is trouble for SMM + A. Since mathematics isn't truth apt on Arealism, SMM + A violates TC. And, as we've already seen, violating TC leads to violating IC. So, SMM + A fails to be a satisfactory naturalistic account of mathematics.

In response, Maddy might say that holding SMM + A responsible for satisfying TC (and so IC) is illegitimate. After all, she's explicit that, from the vantage point of Arealism, set theory (and so mathematics generally) isn't in the business of producing truths. Its aims are purely instrumental. While Maddy can certainly make this move, I don't see that it ultimately helps her. For one thing, it opens her up to the charge that she's abandoned the original project and replaced it with something else. The project of naturalizing mathematics is the project of giving a naturalistic account of mathematical *knowledge*, an account broadly consonant with our best science epistemically and metaphysically. But claiming that mathematics can satisfy its aims without being truth apt is to abandon this project in favor of another – in this case, giving a naturalistic account of mathematical *practice*. This might be a valuable and interesting project, but it's not the one in which we thought we were engaged.

Unlike SMM + A, SMM + TR at least appears to violate MC. SMM + TR endorses the metaphysics of set theory in its entirety. This is a metaphysics not only of abstracta, but of abstracta much of which is so far removed from experience that it begins to acquire an air of supranaturalness. Notice that not all abstracta are like this. Word types,[24] for example, while abstract, bear a reasonably tight connection to experience (viz., experience of word

24 Intuitively, a word type is what multiple copies – tokens – of a word have in common. So, for instance, in the sentence 'All animals are equal but some animals are more equal than others' there are two tokens (each occurrence is a token) of the word 'animals', and they are of the same type, in that they are both the word 'animals'. Similarly for the word 'equal'.

tokens). Similarly for artworks such as string quartets and plays.[25] Even impure sets, sets whose members are concrete, are connected to experience via their members. In all of set theory, however, there are plenty and more pure sets that are hopelessly disconnected from not only actual experience but seemingly from any possible experience. These sets might strike us as even more supranatural than many deities, which are alleged to have quasi-causal influence on the sensible world.

Here again, Maddy might contend that we're making an illegitimate demand on her view. She might argue that SMM + TR is supposed to be free of the sciences, and that demanding it honor MC ties it to the sciences by tying acceptable metaphysics to the sciences. In reply to this, I say that what's of primary importance in MC isn't science but the rejection of supranatural entities. Historically, a driving aim of naturalism has been to account for knowledge on the basis of experience, without appeal to anything that goes beyond experience, including the supranatural. It's this rejection of the supranatural that's of primary importance in MC. As it happens, the best way we have of demarcating the supranatural is by appeal to our best science. This is why science shows up in MC.

Another way Maddy might respond is by claiming that SMM + TR doesn't endorse any particular metaphysics; it endorses the truth (falsity) of set-theoretic sentences, but no specific metaphysics is presupposed or implied to go with this endorsement.[26] This would seem to secure TC and IC, as well as MC, for SMM + TR. But this response has its own problems.

For one thing, the conception of mathematical truth that emerges from Thin Realism strongly resembles an epistemic conception of truth. Epistemic conceptions of truth hold that what's true in a field of inquiry depends in a deep way on the ways we acquire knowledge in that field; we don't *identify* truths through our epistemic activities, rather we *make* things true through those activities. Thin Realism appears to make being a true set-theoretic belief just a matter of being a member of the class of set-theoretic beliefs regulated by the methods (i.e., the epistemic norms) of set-theoretic practice. This is problematic, if only because Maddy herself finds epistemic conceptions of truth lacking.[27] Moreover, such conceptions of truth are *prima facie* inadequate for an explanation of the reliability of set-theoretic practice methods. But one of the central questions supposed to be answered by Thin Realism is just that: Why are set-theoretic methods reliable indicators of truth? The answer on offer here seems (at best) to be that they just are.

This reliability issue leads to a second problem.[28] According to Thin Realism, set-theoretic methods are reliable means to set-theoretic truth because sets just are the sort of thing that the methods of set theory reliably tell us about. A natural question to raise in the face of a claim like this is: What's so special about set theory (mathematics)? There must be something, otherwise practices that we (including Maddy) certainly don't want to put on epistemic par with mathematics could avail themselves of the same sort of argument. For

25 For instance, each performance of Shostakovich's String Quartet No. 8 in C Minor or of *Hamlet* is a token of the relevant work. But every performance of Shostakovich's String Quartet No. 8 in C Minor is of the same type in that each is a performance of the same composition, and every performance of *Hamlet* is of the same type in that each is a performance of the same play.

26 Jody Azzouni's (1994; 2004; 2010) work on truth without metaphysical commitment might be helpful. There's no room to consider this here, but note that since Azzouni is both an antirealist and an antinaturalist about mathematics, it's unlikely Maddy would welcome help from this quarter (see Maddy 2007, § IV.5).

27 See Maddy (2007, § I.7).

28 Cf. my notion of *dissidence* and the discussion in Roland (2007, § 5).

example, certain theists might argue along the same lines that the methods of Christian theology are a reliable means to religious beliefs, including beliefs concerning God and its nature. Such an argument shouldn't be taken seriously and wouldn't be by Maddy. So how is set theory different in such a way that we should take the corresponding argument seriously in its case?

This worry is closely related to one raised by Gideon Rosen (1999). Rosen calls a practice *authoritative* "if, whenever we have reason to accept a statement given the proximate goal of the practice, we thereby have reasons to believe that it is true" (1999, 471). The Authority Problem is the problem of being able to tell authoritative practices from nonauthoritative practices in some principled way. Our problem is a localized version of the Authority Problem: the problem of saying why SMM + TR is authoritative while Christian theology plus a suitably adapted version of Thin Realism isn't.[29]

Maddy understands that this sort of problem arises for her view: Once we allow that mathematics uses methods distinct from those of the natural sciences, what separates mathematics from other practices, such as theology and astrology, which also use methods distinct from those of the natural sciences and which we rightfully reject? Her answer:

> Mathematics is used in science, so the [Second Philosopher's] scientific study of science must include an account of how its methods work and how the theories so generated manage to contribute as they do to scientific knowledge. Astrology and theology are not used in science – indeed, in some versions they contradict science – so the [Second Philosopher] needs only to approach them sociologically or psychologically. (Maddy 2005b, 449)

It is certainly correct to say that mathematics plays a role in science that neither astrology nor theology does; the usefulness of mathematics to science is beyond question. But does this response really tell us why set theory is authoritative and Christian theology isn't? Not obviously.

This response identifies a feature that sets mathematics apart from other nonscientific practices, such as astrology and theology, viz., usefulness to natural science. However, just setting mathematics apart from the other nonscientific practices is not enough to solve the Authority Problem. We can motivate setting mathematics apart from the other nonscientific practices by noting that it is the most notation-heavy among nonscientific practices, but this does not tell us why we should think it is authoritative. Mathematics is certainly special in its usefulness to the sciences. However, absent an argument that usefulness to science is evidence for truth, being special in being useful to science is simply irrelevant to the Authority Problem. Moreover, absent a solution to the Authority Problem for SMM + TR, this view has little claim to satisfying TC and consequently little claim to satisfying IC as well.

Conclusion

In this chapter, I've argued that neither of the two most influential approaches to naturalizing mathematics is able to satisfy all of a certain set of constraints. These constraints are based on central motivations for naturalism generally (EC and MC) or on features of the naturalistic project broadly speaking (TC and IC). In the former case, the constraints ensure

29 Jill Dieterle (1999) raises essentially the same problem concerning astrology for Maddy.

that a naturalistic view bears a substantial connection to the methods and findings of our best science. Satisfying these constraints importantly helps to protect us against inappropriately giving weight to a position in virtue of its being identified as naturalistic. In the latter case, the constraints ensure that the antiskeptical and antirevisionary attitudes of naturalism are respected. If we can naturalize an epistemically successful field of inquiry only by doing violence either to the pool of knowledge of that field or to the ordinary methods of that field, the very methods that have filled that pool of knowledge, then we can't really naturalize *that* field of inquiry. At best we can naturalize some impoverished cousin of that field.

We have seen that both indispensability-based approaches and Maddy's Second Philosophical approach to naturalizing mathematics are unable to satisfy all of the relevant constraints. The problems they have with satisfying the constraints vary somewhat according to the specifics of each approach. There is, however, a sort of overarching problem that I think comes out in considering these approaches and which I'd like to highlight: In order to satisfy TC and IC, a proposal for naturalizing mathematics must capture at least all of classical mathematics; that is, at least all of ZFC. But in doing this, and especially in such a way that it's clear TC is satisfied, one will be hard pressed to satisfy MC. One can back off on the metaphysics in an attempt to satisfy MC. We saw some possible such strategies from both the indispensability-based and Second Philosophical approaches. But in one way or another, these all ended up in trouble with IC. I suspect that something like this holds quite generally, that any attempt to naturalize mathematics will run afoul of at least one of EC, MC, TC, or IC, with the most likely point of failure involving some combination of MC, TC, and IC. If I'm right in this suspicion, then there is no satisfactory way to naturalize mathematics.

Acknowledgments

Thanks to Kelly James Clark for helpful comments.

References

Azzouni, J. (1994). *Metaphysical Myths, Mathematical Practice*. Cambridge: Cambridge University Press.

Azzouni, J. (2004). *Deflating Existential Consequence*. Oxford: Oxford University Press.

Azzouni, J. (2010). *Talking About Nothing*. Oxford: Oxford University Press.

Colyvan, M. (2001). *The Indispensability of Mathematics*. Oxford: Oxford University Press.

Colyvan, M. (2007). Mathematical Recreation versus Mathematical Knowledge. In *Mathematical Knowledge*, edited by Mary Leng, Michael Potter, and Alexander Paseau, pp. 109–122. Oxford: Oxford University Press.

Dieterle, J. (1999). Mathematical, Astrological, and Theological Naturalism. *Philosophia Mathematica* 7: 129–135.

Enderton, H. (1977). *Elements of Set Theory*. San Diego: Academic Press.

Feferman, S. (1988). Weyl Vindicated. In *In Light of Logic*, pp. 249–283. Oxford: Oxford University Press. Includes 1996 postscript.

Feferman, S. (1992). Why a Little Bit Goes a Long Way: Logical Foundations of Scientifically Applicable Mathematics. *Proceedings of the Philosophy of Science Association* 2: 442–455 (with corrections).

Feferman, S. (1998). *In Light of Logic*. Oxford: Oxford University Press.
Feferman, S. (2005). Predicativity. In *The Oxford Handbook of Philosophy of Mathematics and Logic*, edited by Stewart Shapiro, pp. 590–624. Oxford: Oxford University Press.
Feferman, S. and Jäger, G. (1993). Systems of Explicit Mathematics with Non-Constructive μ-Operator, Part I. *Annals of Pure and Applied Logic* 65: 243–263.
Feferman, S. and Jäger, G. (1996). Systems of Explicit Mathematics with Non-Constructive μ-Operator, Part II. *Annals of Pure and Applied Logic* 79: 37–52.
Kanamori, A. (1994). *The Higher Infinite.* Perspectives in Mathematical Logic. Berlin: Springer-Verlag.
Kornblith, H. (1994). Naturalism: Both Metaphysical and Epistemological. In *Midwest Studies in Philosophy,* vol. XIX: *Philosophical Naturalism*, edited by Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, pp. 39–52. Notre Dame: University of Notre Dame Press.
Kunen, K. (1980). *Set Theory*. Dordrecht: North-Holland.
Maddy, P.J. (1997). *Naturalism in Mathematics*. Oxford: Clarendon Press.
Maddy, P.J. (2005a). Mathematical Existence. *Bulletin of Symbolic Logic* 11(3): 351–376.
Maddy, P.J. (2005b). Three Forms of Naturalism. In *Oxford Handbook of Philosophy of Logic and Mathematics*, edited by Stewart Shapiro, pp. 437–459. Oxford: Oxford University Press.
Maddy, P.J. (2007). *Second Philosophy*. Oxford: Oxford University Press.
Maddy, P.J. (2011). *Defending the Axioms: On the Philosophical Foundations of Set Theory*. Oxford: Oxford University Press.
Putnam, H. (1971). Philosophy of Logic. In *Mathematics, Matter, and Method*, 2nd edn. Cambridge: Cambridge University Press.
Putnam, H. (1975). What is Mathematical Truth? In *Mathematics, Matter, and Method*, 2nd edn. Cambridge: Cambridge University Press.
Quine, W.V.O. (1963). Carnap and Logical Truth. In *The Ways of Paradox*, revised and enlarged edn. Cambridge, MA: Harvard University Press.
Quine, W.V.O. (1981). Success and Limits of Mathematization. In *Theories and Things*. Cambridge, MA: Harvard University Press.
Quine, W.V.O. (1998). Reply to Charles Parsons. In *The Library of Living Philosophers,* vol. XVIII: *The Philosophy of W.V. Quine*, expanded and revised edn., edited by Lewis. E. Hahn and Paul A. Schilp, pp. 398–403. La Salle: Open Court.
Roland, J.W. (2007). Maddy and Mathematics: Naturalism or Not. *British Journal for the Philosophy of Science* 58: 423–450.
Roland, J.W. (2008). Kitcher, Mathematics, and Naturalism. *Australasian Journal of Philosophy* 86: 481–497.
Roland, J.W. (2009). On Naturalizing the Epistemology of Mathematics. *Pacific Philosophical Quarterly* 90: 63–97.
Roland, J.W. (2013a). On Naturalism in the Quinean Tradition. In *Philosophical Methodology: The Armchair of the Laboratory*, edited by Matthew C. Haug, pp. 43–61. Abingdon: Routledge.
Roland, J.W. (2013b). Review of *Defending the Axioms: On the Philosophical Foundations of Set Theory. Australasian Journal of Philosophy* 91: 809–812.
Rosen, G. (1999). Review of *Naturalism in Mathematics. British Journal for the Philosophy of Science* 50: 467–474.
Rosenberg, A. (1996). A Field Guide to Recent Species of Naturalism. *British Journal for the Philosophy of Science* 47: 1–29.
Stroud, B. (1996). The Charm of Naturalism. In *Proceedings and Addresses of the American Philosophical Association* 2: 43–55.
Weyl, H. (1918/1994). *The Continuum: A Critical Examination of the Foundation of Analysis*, translated by Stephen Pollard and Thomas Bole. New York: Dover.
Ye, F. (2000). Strict Constructivism and the Philosophy of Mathematics. PhD thesis, Princeton University.

22

Naturalism and Free Will

NEIL LEVY

Varieties of Naturalism

Almost every philosopher working in the analytic tradition accepts some kind of constraint from science. At minimum, analytic philosophers hold that philosophical theories must be *consistent* with our best science. Given that science is our paradigm of a successful epistemic activity, acceptance of this minimal constraint seems almost obligatory; when our theories conflict with settled science, we ought to rethink them.

Naturalists accept stronger constraints on theorizing. However, because "naturalism" is far from a well-defined view, just how individual philosophers understand these constraints differs greatly from thinker to thinker. Naturalism regarding ontological commitments is more widely accepted than methodological naturalism (at least when methodological naturalism is broadly construed); that is, philosophers are typically more concerned with whether theories postulate only entities acceptable to science than with whether theory construction uses only the resources that scientists utilize. It is on ontological naturalism that I will concentrate here.

Views that can call themselves ontologically naturalist, because they postulate only entities consistent with settled science, come in a variety of strengths. Weak ontological naturalism requires only that our theories are consistent with settled science; a view is weakly naturalist if it does not postulate anything supernatural. A stronger version of naturalism demands that the entities and powers cited actually play the kinds of roles that scientists hold them to play. Weaker forms of naturalism, that is, allow themselves a greater degree of latitude with regard to how the properties and powers they invoke are put to work. They make or commit themselves to claims regarding the future evolution of science, for instance, not by claiming that science will discover anything fundamental it does not yet know, but typically by holding that it will answer some currently unsettled question in one way rather than another. Strong ontological naturalists believe that it is more fruitful to accept the current state of science, and not have essential elements of theorizing turn on bets about how things will turn out; not only do they utilize only scientifically respectable properties and powers in their theories, but they defer to science with regard to detailed claims (say, about the architecture of the mind). A mid-strength view avoids bets on the

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

future development of science, by remaining neutral on what will happen, whereas weaker views may allow for speculation about how things will play out as science develops.[1]

Theorists about free will run the entire gamut of positions with regard to science, from those who invoke extrascientific postulates to those who allow themselves a good deal of speculation with regard to what science might one day discover to those who invoke only what is currently known or very confidently believed by mainstream scientists. Given the current primary focus of the free will debate on the compatibility or otherwise of freedom with causal determinism, the science that is at the center of these debates is physics, though because the implementation details matter, the issues also concern the nature of physical processes in the brain. Focusing on physics allows for a neat mapping of positions on free will on to views concerning naturalism: compatibilism with regard to free will and causal determinism predicts acceptance of a stronger versions of naturalism, whereas libertarianism predicts a relatively weak naturalism or the rejection of naturalism. I shall argue, however, that if we focus on different sciences, and especially cognitive neuroscience, this neat mapping breaks down somewhat: compatibilists come to seem committed to much more controversial, and less strongly naturalist, positions.

Free Will and Determinism

The debate over free will has at its center the compatibility question: Is free will compatible with causal determinism? By "free will" is meant some power of agents to act in such a manner that their actions are deeply attributable to them or (perhaps equivalently) that they exercise a distinctive kind of control over their actions. Most theorists understand free will as a power the exercise of which is required for direct moral responsibility for an action; the kind of control at issue is whatever kind of control is required for moral responsibility.[2] "Causal determinism" is a hypothesis concerning fundamental physics: causal determinism is true if the laws of nature determine how events unfold. If causal determinism is true, then given a state of the universe at a time, there is only one possible future state of that universe at the next instant and at all succeeding instants.

For most theorists, the debate over free will therefore has at its center the question whether we are morally responsible for our actions given that states of affairs over which we exercised no control (because they precede our very existence) are causally sufficient for our performing these actions. There are powerful arguments that are alleged to show that free will is incompatible with causal determinism. The *consequence argument*, in particular,

1 Super-strength naturalism not only invokes only properties and entities that play the same role in philosophical theorizing as they do in scientific theorizing, but also looks to model philosophical theories on the details of the science. Super-strength naturalism is more common in the philosophy of mind than in the philosophy of free will, but it is a growing area of research. See Mele (2009), Caruso (2012), and Levy (2014) for examples.

2 For much of the history of the free will debate, there was something approaching a consensus that this power consisted in, or at least entailed, the ability to do otherwise; compatibilists offered conditional analyses of this ability (roughly, an agent has the ability to do otherwise if she would do otherwise were she to decide to do so) and incompatibilists offered "all-in" accounts. Frankfurt (1969) shattered this consensus, by demonstrating to the satisfaction of many participants in the debate that moral responsibility does not require an ability to do otherwise at all. If free will is the control condition on moral responsibility, but moral responsibility does not require the ability to do otherwise, free will does not require any such ability. For further discussion of this knotty issue, see the essays collected in McKenna and Widerker (2003).

has been at the center of many debates. Roughly, the consequence argument holds that since no one has any power over the laws of nature or the states of affairs existing prior to their own existence, and no one has any power over the fact that these states of affairs and laws entail exactly how they will subsequently act, no one has any power over how they act (van Inwagen 1983). We do not act freely because our actions are the consequences of states and processes over which we lack any control.

Responses to arguments for the incompatibility of free will and causal determinism fall into three classes. *Libertarians* accept that these arguments are sound, but deny that determinism is true. *Incompatibilist skeptics* accept that these arguments are sound and also accept that determinism is true, and therefore deny the existence of free will. *Compatibilists* (usually) take no stand on whether determinism is true, but deny that arguments for the incompatibility of free will and determinism are sound.

Given this way of carving up the debate, it is easy to see that compatibilism, at least in its contemporary version, which takes no stand on whether determinism is true, is least committed to controversial scientific claims. However physics unfolds, whether an indeterministic or a deterministic account of quantum mechanics wins out in the end, compatibilists will be untroubled (only so-called Hobartian compatibilists – named after their pseudonymous best-known defender (Hobart 1934) – who hold that free will actually *requires* the truth of determinism, need worry about the future development of science, but Hobartian compatibilism is not a view to which any contemporary philosopher seems to subscribe). One prominent (semi)compatibilist, John Martin Fischer, has argued repeatedly that this lack of commitment to a prediction about the unfolding of future science is a distinct advantage of views like his, because "our basic status as distinctively free and morally responsible agents should not depend on the arcane ruminations – and deliverances – of the theoretical physicists and cosmologists" (Fischer 2006, 5). Compatibilists typically not only postulate no properties or powers that conflict with those identified by settled physics, but take no stand on how debates currently underway within physics are likely to unfold. To that extent, they can claim the mantle of mid-strength naturalists, who avoid bets on the future development of fundamental physics.

Obviously, libertarians do need to take a stand on controversial questions within physics: since they maintain that free will is incompatible with causal determinism, but that we have free will, they are committed to the denial of causal determinism (incompatibilist skeptics, by contrast, are committed to the truth of causal determinism; it is worth noting that many contemporary skeptics avoid this commitment by arguing that free will is compatible neither with determinism nor with indeterminism). Some libertarian naturalists could in fact invoke the rough taxonomy of naturalism we have previously discussed to argue that their position is strongly naturalist, as compared to mid-strength compatibilism. Whereas compatibilists take no stand on how physics will unfold in the future, these libertarians invoke (what they take to be) the current state of science in their theories. They do not merely take a bet about how things will go; they make a claim about the state of science as it is (see the essays collected in Suarez and Adams 2013 for some dramatic examples).

However, the claim to the mantle of strong naturalism by libertarians is somewhat shaky. It might be true that inasmuch as libertarians hold that fundamental physics is more likely to be indeterministic than not, they can invoke support from mainstream science (though I think it would be a stretch to say this question counts as settled). However, there is more to libertarianism than the claim that physics is indeterministic. Libertarians require, at minimum, that the indeterminism be appropriately located – in the mechanisms involved

in decision-making or action production – and that the indeterminism is neither confined to very low-level processes alone, and cancelled out at the level of action and decision, nor so extensive as to undermine control. The latter seems a safe bet, but the former is controversial: it is an open question whether indeterminacies in relatively massive systems like neural networks are likely to be cancelled out or not (Hodgson 2011). Hodgson (2012), after a detailed discussion of the issue, concludes that, at best, contemporary quantum mechanics leaves "room" for genuine and significant indeterminism of the appropriate kind in the brain; hardly a ringing endorsement (which doesn't prevent Hodgson from building his own theory of libertarian free will on similar premises). Clearly, commitment to such a view requires going beyond the evidence of settled science and into speculation.

Further, some versions of libertarianism require even more in the way of controversial claims. Event-causal libertarians maintain that free actions are caused, in the same way in which most action theorists maintain that all actions are caused: by prior mental events. Event-causal libertarians simply add the proviso that prior mental events must *indeterministically* cause our actions. Event-causal libertarians therefore require only that indeterminism be of the right kind and in the right place; beyond that, their claims are clearly consistent with settled science. Agent-causal libertarians seem to require powers and properties beyond those invoked by contemporary physics.

Agent-causal libertarians maintain that, in addition to event causation, there is another kind of causation: agent causation (O'Connor 2000; Clarke 2003). Agents, alone in the universe, have the power to initiate causal chains; they are unmoved movers. How mysterious such an agent-causal power would be is very controversial, but it is uncontroversial that no such power is currently postulated by science. Hence, in making the possession of such a power essential to the possession of free will, agent-causal libertarianism abandons strong and mid-strength naturalism for mere consistency with the findings of science, and therefore, at best, weak naturalism.

The agent-causal claim to weak naturalism is, moreover, shaky on other grounds. It is widely held that naturalism (some would say intellectual respectability) depends on consistency with the causal closure of the physical (for dissent from this orthodoxy, see Hodgson 2012; Adams and Suarez 2013). Since all the evidence we have, and all scientific theories, seem to indicate that all effects have sufficient physical causes, the postulation of nonphysical causes seems to commit us to an unacceptable and unparsimonious overdetermination or (worse) parallelism. In postulating an agent-causal power, however, the agent-causal libertarian seems to commit to the existence of a power that overdetermines physical effects in a manner that clashes with the causal closure principle (Caruso 2012).

Consider, for illustration, the agent-causal view put forward (though not, ultimately, accepted) by Randolph Clarke. On his *hybrid* account, there are two nomologically linked causal routes to action. On the first, event-causal, route, actions are caused (indeterministically) by the agent's having reasons, in precisely the manner suggested by proponents of event-causal libertarianism. But, *in addition*, there is an agent-causal route to action. The reasons an agent has for various courses of action will determine the range of options from among which she chooses, and will influence the relative probabilities of the alternatives (Clarke 2003, 136–137). But the agent retains a distinct power to agent-cause the option she selects: if she agent-causes her decision, and only if she agent-causes it, it will be nondeterministically caused by her reasons. Because her action is caused by her reasons, it is chosen rationally; but because *she* causes it, she has a degree or a kind of control over it unavailable on event-causal libertarian accounts. Be that as it may, the fact that there are two independent

causal routes to the action threatens the causal closure principle. To be sure, Clarke claims that the event-causal route is not sufficient on its own: the action is event-caused only if it is also agent-caused. The real question is not whether his view is consistent with the causal closure principle, given the stipulation: it is whether he is entitled to the stipulation. The closure principle seems to entail that event-causation is sufficient for causation; hence, the stipulation itself should be questioned.

So far as physics alone is concerned, therefore, none of the major positions in the free will debate can claim to be strongly naturalist. Insofar as libertarianism can claim to be in accord with the dominant interpretation of quantum mechanics, it might seem able to make that claim, but the commitment of event-causal libertarianism to claims about how science might develop in the future reduces it to mere consistency with settled science. Agent-causal libertarianism may not even be able to claim that much, insofar as it is questionably consistent with the causal closure of the physical. Mainstream compatibilist views avoid commitment to claims about how physics might evolve, and even to taking a stand on what the dominant view in fundamental physics might be. To that extent, such views are clearly mid-strength naturalisms.

Free Will and Sciences other than Physics

The taxonomy of views in the free will debate is based on their positions with regard to the compatibility question. But (*pace* Balaguer 2010) the question whether we possess free will is not simply the question whether causal determinism is true. Causal determinism might be false, and yet human beings lack free will. For a start, the indeterminism must be of the right kind and in the right location (as we have seen); moreover, other scientific claims also bear on the existence of free will, and therefore we must test how consistent with naturalism a theory is by reference to these claims too.

Consider event-causal libertarianism again. Libertarians of this variety cannot simply stipulate that free actions are those actions that are indeterministically caused by mental states. They must give an account of why actions that are so caused are free actions, and giving that account requires them to spell out details of the causal process that are of dubious scientific status. To see this, consider Robert Kane's event-causal libertarianism, which is surely the best-known and most influential such account available.

Kane (1996; 1999) is concerned, as he must be, not merely with showing that some of our actions are indeterministically caused, but also with showing that these actions are not subject to the so-called luck objection, which is the most common objection to libertarianism in all its varieties. The luck objection turns on the following claim: an agent does not exercise freedom-level control over their actions if it is merely lucky that they acted as they did. Philosophers who wield the luck objection argue that if how I act is indeterministically caused by my mental states, then there is some probability that I will act in one way and some probability that I will act in another, inconsistent with the first, but it is not up to me how the indeterministic causal process unfolds. Nothing about me – my desires, beliefs, values, commitments, and so on – settles whether I ϕ or I φ, since all those states of mine might cause either action. But if I do not control whether I ϕ or I φ, then I do not seem responsible for whether I ϕ or I φ. The very locus of free will, the openness of the future, is shot through with luck, and therefore my control and the degree to which my action is attributable to me is undermined. Libertarians therefore need to show how it can be true

that my actions are indeterministically caused *and* that I exercise freedom-level control over how I act.

Kane's solution is to postulate a relatively detailed theory of how free will is implemented in the brain. Roughly, and abstracting from details, Kane suggests that agents act with direct freedom only when they experience a significant degree of conflict over how they ought to act (a conflict triggered by being torn between the demands of morality and those of self-interest, or between prudential considerations and a desire for immediate satisfaction). When we are torn, we sometimes actually try to perform both (or all) the incompatible actions to which we are sufficiently attracted, he claims. Under these conditions, and these conditions alone, we act with direct free will, because only in these conditions is the brain's thermodynamic equilibrium disrupted sufficiently that the local indeterminism which characterizes low-level neural processes is magnified. In these conditions, a chaotic process begins which is genuinely indeterministic (unlike those chaotic processes that have actually been modeled by scientists, which are deterministic), such that it is physically possible that the neural networks corresponding to each alternative between which we are torn wins out. In these circumstances, it is genuinely possible that we ϕ or φ (limiting ourselves, for simplicity, to two options), but – Kane claims – because we are trying to perform both actions, we are responsible no matter which we actually choose. Kane compares the libertarian agent to a man who attempts to shatter a glass table with his fist. It might be genuinely undetermined whether the table breaks, but because he was trying to break it, if he succeeds he is responsible for breaking it. Similarly, Kane claims, it may be undetermined whether I act morally or selfishly, but whichever I do, because I was trying to do it, I am responsible for my action. The luck objection is thereby seen off (Kane 1999).

It is fair to say that few theorists have been convinced by Kane's response to the luck objection (Mele 2006; Levy 2011). The central objection the view faces is that, though it is true that however the agent acts, she performs an action she was trying to perform, it nevertheless seems a matter of luck whether she performs a good action or a bad one, a prudent action or an imprudent one. In any case, the adequacy of Kane's response to the luck objection is not our concern. Rather, we are concerned with assessing its naturalistic credentials. Kane's view is explicitly designed to be compatible with contemporary science. It invokes no properties or powers that are not scientifically respectable. However, it is clearly not strongly or even mid-strength naturalistic: it is not guided in its details by contemporary science. Rather, it helps itself to a good deal of speculation about how things might possibly go.

We have already seen that the view sketched involves a great deal of speculation about the nature, location, and significance of determinism. Here, I want to highlight the ways in which the neuroscience to which Kane is committed is highly speculative. Competition between neural networks is ubiquitous in the brain, and decision-making may well involve competing networks. But there is no evidence that the phenomenology of being torn disrupts the equilibrium of the brain or is associated with more intense competition between networks. Similar remarks might be made about David Hodgson's (2012) libertarianism. Hodgson's theory turns on the alleged capacity of conscious processes to grasp information as unified gestalts and thereby contribute to undetermined and nonformalizable decision-making. But there is little evidence that conscious processes have an advantage over nonconscious in the respects Hodgson identifies (they may, indeed, have a disadvantage in these respects, given their smaller capacity). Further, there is some evidence to suggest that consciousness is the domain of rule-governed processes (Levy 2013a). Given that this is the case, it is more likely that nonconscious processes are governed by processes that are not formalizable than conscious processes.

So far as physics is concerned, compatibilist accounts of free will tend to leave fewer hostages to fortune: because they do not see themselves as needing to address the luck objection, compatibilists do not see a need to put forward speculative accounts of how decision-making might be implemented in agents like us. It is precisely because they can avoid committing themselves that they are often mid-strength naturalist views: neither taking a stand on controversial questions, nor basing their position on how things seem most likely to turn out. However, when we turn from physics to other sciences, compatibilists often turn out to be implicitly committed to controversial claims.

Consider John Martin Fischer's semicompatibilism (compatibilism about moral responsibility, combined with agnosticism about free will; the issues Fischer raise seem common to other compatibilist accounts, so it is appropriate to focus on the admirably detailed way in which he has developed them). Fischer maintains that agents are morally responsible for their actions when they exercise *guidance control* over them, and that they exercise guidance control over actions caused by moderately reasons-responsive mechanisms of their own (Fischer and Ravizza 1998; Fischer 2006). A mechanism is moderately reasons-responsive when it is appropriately receptive and reactive to reasons, including moral reasons. To be receptive to reasons is to be able to recognize them *as* reasons; a mechanism is receptive to reasons, in the kind of way required, if it would recognize a (broad enough) range of reasons to act. A mechanism is reactive to reasons if there is at least one possible world in which, in response to one of the reasons to which it is receptive, the mechanism would actually cause a different action to the one it caused in the actual world. Every element of Fischer's account commits him to empirical claims that may be more controversial than he seems to recognize.

We test to see whether agents exercise guidance control over their actions by asking how they would have acted in counterfactual circumstances. Suppose an agent performs a morally bad act (stealing from an impoverished person, say). It is a necessary condition of their being blameworthy that in counterfactual circumstances that resemble (but are not identical to; identity would build in a commitment to indeterminism) the actual circumstances, they would have acted otherwise, with their action being caused by the very same mechanism upon which they acted in the actual world. If there had been appropriate incentives to act otherwise – say the probability of being caught was raised, or the punishment for being caught was likely to have been greater – would the agent have recognized these reasons *as* reasons to act otherwise? If the answer is yes, then the agent is appropriately receptive to reasons, and we can turn to reactivity. We assess reactivity by asking whether there is at least one possible scenario in which (again holding the actual mechanism fixed) the agent would have done otherwise. Only if they were appropriately receptive *and* reactive is the agent blameworthy.

Despite Fischer's professed aim to avoid having moral responsibility hang on the thread of abstruse ruminations by scientists, the account he offers is committed to a number of controversial and quite possibly false claims. First, he is committed to claims about mechanisms and their individuation. Fischer and Mark Ravizza (with whom Fischer initially worked out the detail of the theory under discussion) do not offer a theory of mechanism individuation. Rather, they rely upon our "relatively clear intuitive judgments" about which mechanisms are involved in which behaviors (Fischer and Ravizza 1998, 40). Sometimes, they allege, the fact that an agent would act otherwise given an incentive is not evidence for moderate reasons-responsiveness, because in the counterfactual scenario the agent acts on a different mechanism; sometimes, presumably, the failure to act otherwise in a counterfactual

scenario is not evidence for lack of reasons-responsiveness, because the mechanism upon which the agent acts is not held fixed. It is, however, a live issue whether our intuitions about the mechanisms involved in behaviors track the actual mechanisms involved; indeed, there is a question whether we can make much neuropsychological sense of mechanism talk in the way Fischer engages in it at all.

We might attempt to identify the mechanisms required with the *modules* postulated by many cognitive scientists; discrete and dissociable processing units, which may be informationally encapsulated from one another and therefore insensitive to information that falls outside their own proprietary domain. Cognitive science has identified many candidates for mechanisms that subserve aspects of practical reasoning and behavior, from basic perpetual systems to higher-level categorization systems, language production and comprehension systems, and possible domain-general systems like Sperber's (1996) formal logic module. There is evidence that behavior is sometimes driven by such mechanisms. Phobias, for instance, might be paradigms of such behaviors: in such cases, because the mechanism that generates what Gendler (2008) calls a representational, affective, and behavioral complex is insensitive to domain-general information – it is not modulated by the personal-level belief that the spider is not dangerous (or indeed is made of rubber) – the behavior is driven by a discrete mechanism. But few of our behaviors seem to be like this: most are driven by the interaction between a large number of mechanisms. Testing for reasons-responsiveness by holding just the actual set of mechanisms fixed risks making the theory extensionally equivalent to incompatibilism: agents will not pass the test in any (or, at any rate, in more than a tiny number of) counterfactual situations, because normal behavior is typically driven by mechanism *switching*.

Fischer may think that he avoids commitment to any controversies in cognitive science by individuating mechanisms by intuition. But if we are interested in the truth about agents' capacities, we cannot rest content with individuating mechanisms by intuition. Presumably, we cannot rightfully blame agents for their wrongful actions on the basis that though they lacked the capacity to act as they ought, we have the intuition that they suffered no such lack.

Turn now from the question of which mechanism agents act upon to the question of their responsiveness to reasons. This question confronts all compatibilists, whether or not they accept Fischer's claims about mechanisms, to the extent they assess the voluntariness of behavior by asking how the agent would have acted given appropriate incentives. To a first approximation, at least, this question is therefore an issue for all compatibilists, from classical compatibilists who advance conditional analyses of the ability to do otherwise (an agent has the capacity to act otherwise just in case she would act otherwise given a sufficient reason to do so), to new dispositionalists who hold that an agent has the capacity to act otherwise just in case she has dispositions to act appropriately (Smith 2003; Vihvelin 2004), to reasons-responsiveness theorists who replace Fischer's talk of mechanisms with talk of the agent and her capacities. All these theorists hold that a necessary condition of holding an agent morally responsible for an action is that she has some capacity or set of capacities that would motivate alternative behaviors given the appropriate incentives.

Again, though, all these theorists rely upon intuition in testing whether the mechanism or agent is reasons-responsive. And again, these intuitions may be misleading. All these theorists are committed to the claim Fischer and Ravizza make explicit, that "reactivity is all of a piece" (1998: 73); that is, if a mechanism is reactive to *some* incentive to do otherwise, than the mechanism is reactive to *any* intuitively similar or intuitively stronger incentive to

do otherwise. Thus, if an agent who commits a wrongful act would refrain were she threatened with 1 week's imprisonment, then she would refrain were she threatened with 2 weeks', or 3 years', or 2 decades' imprisonment; thus, if an agent would agree to paint my house for $100, she would do it for $101, $425, or $5800. Thus, too, if an agent would refrain from committing a wrongful act if offered $20 to do so, she would refrain if threatened with 6 months' detention.

The problem is, some of these claims may be false. Reactivity to incentives may be more local and more patchy than we tend to think. Consider, for instance, some of the evidence Gene Heyman (2009) cites for the conclusion that addiction is "a disorder of choice." Heyman bases this conclusion on evidence that addicts respond to incentives, coupled with the implicit assumption that reactivity is all of a piece. Thus, from evidence that addicts often give up their addiction in response to ordinary demands of life (the acquisition of responsibility for a child, say), or in response to very modest financial incentives (such as the low-value vouchers offered by Steve Higgins et al. 1994 in experiments on cocaine addicts), Heyman concludes that addicts can always act otherwise than they do. But the inference is too quick, because the assumption that reactivity is all of a piece is probably false.

Indeed, Higgins' own work demonstrates the point. As Heyman notes, the value of the vouchers Higgins offered his subjects was always very much less than the subjects were already spending on cocaine. If the financial incentive was motivating, and reactivity is all of a piece, then one would expect the vouchers to have a motivational force very much smaller than the motivational force of the expenditure required to acquire cocaine. Clearly, we cannot simply read agents' reactivity off from their responsiveness to a small range of incentives. This remains true not only for addicts, but for all of us.

There is plentiful evidence that discount, delay, and uncertainty have unexpected effects on the motivational force of incentives, and that therefore these effects are counter-intuitive (Ross et al. 1998). There is also evidence that positive and negative incentives do not line up in the way we expect. The risk of the loss of $10 is in some circumstances very much more motivating than the potential of gaining $10, even controlling for how badly the subject needs the money. As a consequence, framing the same outcome as a loss leads subjects to respond more strongly – be more highly motivated – than if it is framed as a forgone gain (Tversky and Kahneman 1981). However, in other circumstances, for reasons that are ill understood, we respond *more* strongly to incentives than to equivalent punishments (Ross et al. 1998; Levy 2013b). Work in this field is rapidly evolving, but we know enough to be confident about at least this much: we cannot always rely on our intuitions in attempting to assess whether an agent we know to be generally rational would react to a particular incentive. Even in our own case, I suspect, such intuitive judgments will often be wrong.

When we map the free will debate by reference to the traditional focus – the question of determinism – and ask how the rival theories fare with regard to how scientifically plausible they are and how much speculation about the future course of science they engage in, we get a neat continuum, from compatibilism through to libertarianism. Compatibilism seems to have better naturalist credentials and leave fewer hostages to scientific fortune, qualifying as a mid-strength naturalism. When we broaden the scope to other sciences, and especially the sciences of the mind, this advantage of compatibilism weakens, and the picture becomes much muddier. Speculation seems rampant on all sides. Mere consistency with settled science seems the best any mainstream view can claim for itself.

Is Naturalism a Virtue?

Let us turn now from the question of which side can claim better or stronger naturalist credentials to whether any side should be interested in claiming such credentials.

There are many possible ways of doing philosophy, and many possible roles that philosophical reflection can play. Philosophical theories can be assessed for their consistency with contemporary science – by which measure none of the best-known theories fares outstandingly well – but they can also be assessed by different standards. In evolutionary biology, theorists sometimes offer "how possibly" stories. They seek to show how a particular trait *might* have evolved, rather than to argue that the story offered must be true, or even that it is an excellent candidate for truth. Even from a scientific point of view, this exercise is useful: it generates testable hypotheses. Offering a "how possibly" story requires consistency with science, but it does not require strong naturalism: the details of the story need not match up closely with what we already know.

Philosophical theories can be offered in this spirit, and count as naturalistically respectable. For instance, Kane's speculations on how the brain might work to generate the physical possibility of conflicting actions might be used to generate hypotheses. However, insofar as claiming naturalistic respectability is important to philosophers, they must take care to avoid a vice sometimes exhibited by those evolutionary biologists who offer "how possibly" stories: taking the story to be not merely a hypothesis but an explanation, to be defended against all objections and at almost any intellectual cost (or alternatively, utilizing it as a premise in an argument to a controversial conclusion without recognizing that the uncertainties with regard to the premise should lower our confidence in the conclusion). There are signs that philosophers like Kane and Hodgson do not offer their theories in this kind of spirit at all. Hodgson, for instance, moves from demonstrating, reasonably convincingly, that there is "room" in contemporary physics and neuroscience for significant indeterminism to asserting confidently that we have libertarian free will.

However, perhaps philosophers like Kane and Hodgson do not advance their theories in the kind of spirit that motivates this criticism. Perhaps naturalistic credentials are less important to them than they would have to be for the criticism to have any bite, by *their* lights. Many philosophers implicitly or explicitly claim that our commitment to agency, responsibility, and freedom should trump our commitment to science. Peter van Inwagen (1983), whose formulation of the consequence argument has been so influential in motivating contemporary incompatibilism, has said that were he to be convinced of the truth of determinism, he would immediately convert to compatibilism; for him, our possession of free will is a fixed point, which cannot be overcome on the basis of scientific discovery (there is experimental evidence that suggests van Inwagen's position is also held by many ordinary people, who would adjust their position on the compatibility of moral responsibility and causal determinism depending on whether the actual world turns out to be deterministic or not; see Roskies and Nichols 2008). Despite holding his views about freedom fixed, and thereby insulating them from scientific discovery, there is one way in which van Inwagen's view nevertheless defers to science: if science turns out to go one way rather than another, he will adjust his philosophical stance. Other philosophers have gone rather further.

Helen Steward (2012), for instance, asserts that since agency itself (and not merely freedom) is incompatible with causal determinism, we ought to conclude that determinism is false. Steward maintains some naturalistic credentials by attempting to withdraw the question of determinism from the purview of science: just as van Inwagen sees free will as

an issue within metaphysics, upon which science bears but about which it is not equipped to pronounce, so Steward sees determinism as at least partially an issue in metaphysics. Steward accepts that facts about agents and institutions supervene, at a time, at any rate, on physical facts. But supervenience – by itself – does not entail that there is a one-way causal relationship between physical facts and facts about agency and institutions. Physical facts may depend, in part, on facts about agents, consistent with supervenience. Since Steward believes that agency requires the falsity of determinism, and that it is inconceivable that agency does not exist, she thinks we have good evidence for the falsity of determinism. Like van Inwagen, Steward says she would rethink her incompatibilism were she to discover that determinism is true (while she denies that the question whether determinism is true is a question that can only be answered by physics, she does not deny that physics may one day answer the question). But whereas van Inwagen does not appear to think that his incompatibilist arguments provide *evidence* that indeterminism is false, Steward believes that her arguments for incompatibilism, together with her arguments for the claim that determinism is as much a metaphysical as a physical question, do provide (defeasible) evidence for the falsity of determinism (somewhat analogously, Gisin 2013 has argued against the many-worlds interpretation of quantum mechanics on the grounds that it would rule out libertarian free will).

There seems to be something a little odd about Steward's claims here. She claims (1) that we are agents at all only if determinism is false and (2) that there need not be a one-way causal relationship between physical facts and facts about agency. But if is true that facts about agency may settle what physical facts obtain, then it seems false that the truth of determinism makes a difference to whether we are agents. Determinism, which is a thesis about physical facts, threatens our agency only if the causal arrow runs from the physical facts to facts about agency and not vice versa. At any rate, we can leave this question aside, to ask instead about the naturalistic credentials of Steward's view.

Mere consistency with the claim that facts about agency supervene on physical facts is not sufficient to establish that a view is naturalistic, in any genuine sense. As Papineau (2007) points out, a very wide variety of properties with dubious naturalistic credentials can satisfy the supervenience condition; it is quite possible to accept (as did Moore with regard to moral properties) that a particular property supervenes on physical properties and to be explicitly antinaturalist about this property. I suspect this charge won't worry Steward; if her views are best described as antinaturalist, she will accept that description. However, there are good reasons to want something stronger than mere supervenience: without something stronger, causal closure seems threatened. Two centuries of work on the brain seem to indicate that neural events are causally sufficient for thought and action. If thought and action *merely* supervene on such events, because there is no one-way causal link between the second and the first, it seems that we are required either to postulate nonphysical forces, contrary to the closure principle, or overdetermination, which seems absurd. Since the closure principle appears well established, these costs seem too great to pay.

Perhaps Steward could reply by saying that however well the truth of the closure principle seems to be established, the fact of agency is better established, so agency should trump closure. The cogency of that reply will depend on the strength of her argument that agency is threatened by something stronger than mere supervenience. It is important to note that it is open to the incompatibilist to accept that agency is incompatible with causal determinism without accepting that it is incompatible with anything stronger than mere supervenience; for that reason, Steward's apparent commitment to denying causal closure does not extend

to other libertarian views concerned with agency (it goes without saying that it does not extend to other libertarian views that require indeterminism for freedom, but not for agency). Agent-causal views seem to have problems of their own with causal closure, but event-causal views have no such problems. The closure principle is therefore a weak constraint on theories regarding free will.

Apart from concerns about agency, there is one consideration that might motivate resistance to naturalism, of any variety, in the free will debate. It is this: free will may plausibly be seen as at least partially a normative question, and normative properties seem to many to be more resistant to naturalization than are non-normative. The free will debate at least partially concerns normative properties for all those philosophers who understand free will as a necessary condition for moral responsibility (e.g., McKenna 2004; O'Connor 2005; Nahmias et al. 2008). Of course, some philosophers hold that moral properties can be naturalized, but it is worth remarking that extant naturalistic analyses of normative properties seem to succeed far better with concepts like "good" than those like "desert." Some philosophers deny that there is any necessary connection between free will and moral responsibility. It is a difficult question whether agency ought to be considered a partially normative property; if it is, it is surely one of the normative properties that is most easily naturalized.

The prospects for a fully naturalistic account of free will are therefore tied up with the prospects for fully naturalistic accounts of other properties. If agency cannot be naturalized, the prospects seem dim; fortunately, relatively few philosophers doubt that agency can be naturalized. It is more controversial whether "responsibility" and "desert" can be naturalized; whether a failure to naturalize them would have implications for a strong naturalistic account of free will depends on one's account of free will. Some apparently naturalism-resistant properties like "consciousness" may also play a role in some accounts of free will central enough to create problems for a strong naturalism (see Caruso 2012 for discussion). At present, no view can claim to be more than mid-strength naturalist: all views take some bets on how the science will turn out. Mid-strength naturalism might be naturalism enough, however: it is certainly naturalism enough to expose the view to naturalistic refutation (we can already see that some of the bets mid-strength naturalist views take are likely to be bad ones). Stronger naturalisms await the further development of science.

Acknowledgments

I am grateful to Kelly Clark for comments on an earlier draft.

References

Adams, P. and Suarez, A. (2013). Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience. In *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*, edited by A. Suarez and P. Adams, pp. 273–290. New York: Springer).

Balaguer, M. (2010). *Free Will as an Open Scientific Problem*. Cambridge, MA: MIT Press.

Caruso, G. (2012). *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lanham: Lexington Books.

Clarke, R. (2003). *Libertarian Accounts of Free Will*. New York: Oxford University Press.

Fischer, J.M. (2006). *My Way: Essays on Moral Responsibility*. Oxford: Oxford University Press.

Fischer, J.M. and Ravizza, M. (1998). *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66: 829–839.

Gendler, T. (2008). Alief and Belief in Action (and Reaction). *Mind & Language* 23: 552–585.

Gisin, N. (2013). Are There Quantum Effects Coming from Outside Space–Time? Nonlocality, Free Will and "No Many-Worlds." In *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*, edited by A. Suarez and P. Adams, pp. 23–29. New York: Springer.

Heyman, G. (2009). *Addiction: A Disorder of Choice*. Cambridge, MA: Harvard University Press.

Higgins S.T., Budney, A.J., Bickel, W.K., Foerg, F.E., Donham, R., and Badger, G.J. (1994). Incentives Improve Outcome in Outpatient Behavioral Treatment of Cocaine Dependence. *Archives of General Psychiatry* 51: 568–576.

Hobart, R.E. (1934). Free Will as Involving Determination and Inconceivable Without It. *Mind* 63: 1–27.

Hodgson, D. (2011). Quantum Physics, Consciousness, and Free Will. In *The Oxford Handbook of Free Will*, 2nd edn., edited by Robert Kane, pp. 57–83.

Hodgson, D. (2012). *Rationality + Consciousness = Free Will*. Oxford: Oxford University Press.

Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.

Kane, R. (1999). Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism. *Journal of Philosophy* 96: 217–240.

Levy, N. (2011). *Hard Luck*. Oxford: Oxford University Press.

Levy, N. (2013a). Critical Notice: *Rationality + Consciousness = Free Will*. *Australasian Journal of Philosophy* 91: 183–192.

Levy, N. (2013b). Punishing the Addict: Reflections on Gene Heyman. In *The Future of Punishment*, edited by Thomas Nadelhoffer, pp. 233–245. Oxford: Oxford University Press.

Levy, N. (2014). *Consciousness and Moral Responsibility*. Oxford: Oxford University Press.

McKenna, M. (2004). Compatibilism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://setis.library.usyd.edu.au/stanford/entries/compatibilism (last accessed July 15, 2015).

McKenna, M. and Widerker, D. (eds.) (2003). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate Press.

Mele, A. (2006). *Free Will and Luck*. New York: Oxford University Press.

Mele, A. (2009). *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.

Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2008). Is Incompatibilism Intuitive? In *Experimental Philosophy*, edited by Joshua Knobe and Shaun Nichols, pp. 81–104. Oxford: Oxford University Press.

O'Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

O'Connor, T. (2005). Free Will. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://setis.library.usyd.edu.au/stanford/entries/freewill/ (last accessed July 15, 2015).

Papineau, D. (2007). Naturalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/naturalism/ (last accessed July 15, 2015).

Roskies, A. and Nichols, S. (2008). Bringing Moral Responsibility Down to Earth. *Journal of Philosophy* 105: 371–388.

Ross, D., Sharp, C., Vuchinich, R.E., and Spurrett, D. (2008). *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling*. Cambridge, MA: MIT Press.

Smith, M. (2003). Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet, pp. 17–38. New York: Oxford University Press.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.
Steward, H. (2012). *A Metaphysics for Freedom*. Oxford: Oxford University Press.
Suarez, A. and Adams, P. (eds.) (2013). *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*. New York: Springer.
Tversky, A. and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science* 211: 453–458.
van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Oxford University Press.
Vihvelin, K. (2004). Free Will Demystified: A Dispositional Account. *Philosophical Topics* 32: 427–450.

23

Free Will and Naturalism: How to Be a Libertarian, and a Naturalist Too

KEVIN TIMPE AND JONATHAN D. JACOBS[1]

Introduction

As pop naturalists tell it, free will is incompatible with naturalism. And apparently many scientists agree. Philosopher Daniel Dennett reports, for example, that he has "learned from discussions with a variety of scientists…[that] free will, in their view, is *obviously* incompatible with naturalism, with determinism, and very likely incoherent against any background, so they cheerfully insist that of course they don't have free will" (2013, 47). Many philosophers, however, disagree (e.g., Mele 2014; Nahmias 2014; Vargas 2014), since compatibilist forms of free will seem amendable to purely naturalist underwriting. There is nevertheless, among philosophers, a near consensus on this: naturalism is certainly incompatible with *libertarian* free will.

We aim to show in this paper that libertarian free will is not incompatible with naturalism. Even the purportedly "spooky" – or "colossally whackadoodle"[2] – libertarian form of free will known as agent causation can be situated within a naturalistic metaphysical framework. Neither of us, though, is a naturalist,[3] and you know what they say about enemies bearing gifts. Still, our goal in this chapter is to offer a framework for an agent-causal account of free will that naturalists should be able to accept.

1 Order of authorship was determined by a coin flip.

2 This description of the view at issue here was relayed by Michael McKenna in personal correspondence.

3 We are both theists. Nevertheless, we think that an account of free will can be given that – *qua* account of the capacities involved in free will – does not require the existence of God. The qualifier "*qua* account of the capacities involved in free will" is needed for two reasons. First, according to traditional theism, God is the "creator of all things, seen and unseen." Thus, if created agents have free will, its existence depends upon God's creating and sustaining it in existence. The same is also true of giraffes – but we don't think an account of the nature of giraffes should make explicit reference to the existence of God. Second, we think that not only does God exist, but He exists necessarily, and so it's not perfectly clear what we should say about free will were God not to exist. However, we also think that not all counterpossibles are only trivially true; see, for instance, Zagzebski (1990).

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

We proceed as follows. We begin by addressing some terminological issues and methodological constraints. We then examine recent claims by naturalists, particular a number of what might be called "pop naturalists," about the nonexistence of free will, and argue that they have failed to justify their conclusions. We then develop a framework for an incompatibilist account of free will that ought to be acceptable to the naturalists (and, for the most part, is also acceptable to the compatibilist). But first, some terminological and methodological issues.

Terminology and Methodological Constraints

For those not familiar with the relevant literature, it may come as a surprise that the meaning of the term "free will" is contentious. Peter van Inwagen, one of the most influential figures in contemporary free will debates, argues that the term "free will" should be defined in terms of the ability to do otherwise[4]. According to van Inwagen (2008, 329), free will just *is* "hav[ing] both the following abilities: the ability to perform that act and the ability to refrain from performing that act. (This entails that we have been in the following position: for something we did do, we were at some point prior to our doing it able to refrain from doing it, able not to do it.)" Yet others (Vargas 2007a; McKenna 2008, 30; Timpe 2012, 9–10) deny this and prefer to define free will in terms of the control condition on moral responsibility; that is, they think that having free will just is controlling your actions in the way required to be morally responsible for those actions.[5]

For present purposes, we stipulate the following definition of free will:

Free will $=_{df}$ the capacity or set of capacities of an agent to control her choices or volitions, the exercise of which is necessary for the agent to be morally responsible for those choices or volitions.[6]

We use "choices or volitions" in the definition rather than "actions" because we think that an agent can freely choose even if she is unable to carry out the action she chose.

Both of us are *incompatibilists*; that is, we think free will is incompatible with causal determinism. In order for human agents to be free and responsible, causal determinism – "the thesis that the past and the laws of nature together determine, at every moment, a unique future" (van Inwagen 2008, 330) – must be false. We will not rehash our arguments for incompatibilism here (though see, for instance, Timpe 2012). Nor will we argue here for the existence of free will,[7] though we will return to some naturalist denials of it in the next section. Rather, we're simply going to assume the conjunction of incompatibilism and the existence of free will, a view called "libertarianism."

One reason for giving an account of free will that is consistent with, even if it doesn't require, the falsity of determinism is as follows. Few contemporary compatibilists think that

4 Actually, this isn't quite correct; what van Inwagen defines is not free will but the "free will thesis." Van Inwagen advises that one "define sentences, not terms" (2008, 328).

5 Others take these two understandings of free will to be coextensive; see, for example, Kane (2001, 17). Yet others think that they can come apart; see, for example, Fischer and Ravizza (1998).

6 For reasons why one of us prefers this approach to defining "free will" over van Inwagen's approach, see Timpe (2012, 11). We take this definition to line up with how Levy is using the term in Chapter 22.

7 For how such arguments might go, see Timpe (2012, esp. ch. 4). In our view, Neil Levy gives one of the most compelling arguments for free will skepticism in Levy (2011). For a partial response by one of us to his challenge, see Timpe (2012, ch. 10).

free will requires the truth of determinism. Instead, they are "free will either way" theorists; that is, on their view, the truth or falsity of determinism is irrelevant to whether or not we are free. Manuel Vargas (2012, 420) has recently given the name "supercompatibilism" to this kind of compatibilist view. John Martin Fischer thinks that a major advantage of contemporary compatibilism is precisely that we can be free whether or not determinism is true. Compatibilism is resilient, according to Fischer, to the discoveries of physicists in a way that libertarianism is not:

> I could certainly imagine waking up some morning to the newspaper headline, "Causal Determinism Is True!" (Most likely this would not be in the *National Enquirer* or even *People* – but perhaps the *New York Times*…) I could imagine reading the article and subsequently (presumably over some time) becoming convinced that causal determinism is true…And I feel confident that this would not, nor should it, change my view of myself and others as (sometimes) free and robustly morally responsible agents…The assumption that we human beings – most of us, at least – are morally responsible agents (at least sometimes) is extremely important and pervasive. In fact, it is hard to imagine human life without it…A compatibilist need not give up this assumption [that we are at least sometimes free and morally responsible], even if he were to wake up to the headline, "Causal Determinism is True!" (and if he were convinced of its truth)…A compatibilist need not "flipflop" in this weird and unappealing way. (2007, 44–47)[8]

In contrast, if a libertarian were to become convinced that determinism were true, she would have to give up either her belief that we are morally responsible beings or her incompatibilism. "One of my main motivations for being a compatibilist," writes Fischer, "is that I don't want our personhood and our moral responsibility, as it were, to hang on a thread, or to be held hostage to the possible scientific discovery that determinism is in fact true" (2000, 323).[9] If the compatibilist's account of free will required determinism to be true, it would be just as susceptible to flipflopping as would the libertarian's view. And for all we currently know, determinism is false. Richard Holton, summarizing the data concerning determinism, says that "the science is complicated and uncertain" (2013, 87); similarly, Randolph Clarke writes that "perhaps the best that can be said…is that…there is no good evidence that determinism is true" (2002, 377).[10] But for supercompatibilists, if determinism turns out to be false, their accounts of the nature of free will should be able to easily accommodate indeterminism.[11] Hence, it seems dialectically acceptable – even preferable – to assume that determinism is false.

In this first section, we also want to lay out some methodological constraints on how we should think about free will. Manuel Vargas introduces what he calls "the standard of naturalistic compatibility": "As I use the phrase, naturalistic compatibility refers to coherence

8 It should be clear that Fischer is thinking specifically of what Vargas calls "supercompatibilism" here. If, contrary to the general thrust of contemporary compatibilists, free will required the truth of determinism, then compatibilism would not enjoy the dialectical advantage that Fischer here attributes to it. For more on the charge of "flipflopping," see Steward (2012, ch. 5).

9 See also Fischer (1999, 129; 2006, 183). For another presentation of this worry, see Vargas (2007b, 141ff.). See also the discussion in Chapter 22 of this volume.

10 Psychologist George Howard claims that "if you want to be a scientist, you had better be a determinist" (2008, 261). The rest of the passage, however, shows that he wrongly assumes that all causation must be deterministic.

11 For an argument that the supercompatibilist ought to join with the libertarian is refuting the "luck objection" for similar reasons, see Timpe (2012, 176).

with the known facts, along with compatibility with two further things: (1) the principle of the causal closure of the physical; and (2) acceptance of a principle of methodological minimalism" (2013, 56). By (2), he means "the idea that additions to our ontology of entities and powers have to be earned in terms of familiar theoretical virtues like explanatory power, prediction, and so on. In concrete terms, what this means for contemporary libertarianism is that it does not (ordinarily) appeal to anything like the spooky powers of old: substance dualism, nominal selves, or a 'god-like' power to be a 'unmoved mover'" (2013, 56ff.). This weak claim parallels what Levy claims is true of analytic philosophy more generally: "At minimum, analytic philosophers hold that philosophical theories must be *consistent* with our best science" (Chapter 22). We're willing to endorse, for present purposes, the standard of naturalistic compatibility and the principle of methodological minimalism. We'll address the causal closure of the physical further on.

Following Levy, we'll concern ourselves primarily with "ontological naturalism," the view that any philosophical view should "postulate only entities acceptable to science" (Chapter 22). As Levy notes, ontological naturalism comes in a variety of strengths. At the very least, the framework we outline in this chapter is an example of what he calls "weak ontological naturalism," since it does "not postulate anything supernatural" (Chapter 22). However, the proposed view may be compatible with stronger forms as well, meeting what Vargas calls "the standard of naturalistic plausibility": "the standard of naturalistic plausibility is satisfied if the account satisfies the standard of naturalistic compatibility and there are truth-relevant considerations that speak in its favor when it postulates requirements that exceed the known facts or the widely accepted ontologies of our current scientific understanding" (2013, 59).[12] Perhaps most importantly, we don't think that the proposal violates this standard in virtue of including substance causation. As we see it, the nature of the causal relata is not settled by physics, and thus the framework's commitment to substance causation doesn't clearly rule out a stronger naturalistic position (Harre 2001). And if all causation is substance causation, then the framework need not appeal to causal powers beyond those invoked by contemporary physics; nor need it be committed to "the existence of a power that overdetermines physical effects in a manner that clashes with the causal closure principle" (Chapter 22).

The view that all causation is substance causation, the details of which we spell out further on, need not entail that causal closure is false.[13] However, assuming that objects like rocks and people are substances, as we are wont to do, it would be inconsistent with *micro*-physical closure: the view that every physical event has an event involving only small, microphysical objects as its cause. But there are reasons to think that naturalism doesn't commit one to microphysical closure. For one, a number of naturalists themselves argue for the falsity of microphysical closure (see, e.g., Dupré 1993; Lowe 2000; Papineau 2001; 2009). Furthermore, given that it is an open scientific question whether microphysical closure is true, we think it would be odd for the truth of naturalism to depend on this kind of closure being true insofar as the truth of the former would then "hang by a thread" in just the way that naturalists like Fischer think other philosophical views ought not. Imagine the following parallel of Fischer's conversation regarding incompatibilism:

12 It's not clear to us that there is a "widely accepted ontology" in our current scientific understanding, but we won't pursue that point here.

13 In Chapter 22, Levy raises a worry about closure if one postulates both event- and agent-causation. However, insofar as the view being developed here only affirms substance causation, we can avoid his concern.

> I could certainly imagine waking up some morning to the newspaper headline, "Microphysical Closure is False!" (Most likely this would not be in the *National Enquirer* or even *People* – but perhaps the *New York Times*…) I could imagine reading the article and subsequently (presumably over some time) becoming convinced that microphysical closure is false…And I feel confident that this would not, nor should it, require me to change my view that naturalism is true.

It would certainly be odd if the discovery that a nonmicroscopic, physical object caused a physical event led anyone to reject *naturalism*. For these reasons, we think naturalism ought not require the truth of microphysical closure. Thus, even though our view in this chapter will be incompatible with the truth of microphysical closure, the view is not for that reason incompatible with naturalism.

Naturalists on Free Will

Vargas thinks it is a "real shame" that "outside of specialists working on free will and moral responsibility, there is sometimes the impulse to dismiss out of hand any form of libertarianism" (2013, 53). We agree, and in this section we briefly consider what a number of prominent naturalists have said about free will and argue that many make claims that are not justified by the philosophical literature on the subject. Once we have set aside these confusions, we turn then in the next section to developing the outlines of a libertarian position that ought to be acceptable to the naturalist.[14]

We begin with a "popular" naturalist critique of free will given by Jerry Coyne, a biologist on the faculty at the University of Chicago. In a series of recent, widely publicized articles, Coyne argues that none of our choices result from "free and conscious decision on our part. There is no freedom of choice, no free will" (2012a). Coyne begins an article in *The Chronicle Review* as follows:

> The term "free will" has so many diverse connotations that I'm obliged to define it before I explain why we don't have it. I construe free will the way I think most people do: At the moment when you have to decide among alternatives, you have free will if you could have chosen otherwise.[15] To put it more technically, if you could rerun the tape of your life up to the moment you make a choice, with every aspect of the universe configured identically, free will means that your choice could have been different. Although we can't really rerun that tape, this sort of free will is ruled out, simply and decisively, by the laws of physics. (2012b)

Coyne's argument proceeds roughly as follows: (1) determinism is true, (2) free will just is the ability to do otherwise, and (3) if determinism is true, everyone lacks the ability to do otherwise. We are not convinced, as is Coyne, that science has shown that (1) is true; he certainly hasn't demonstrated that it is. Coyne's endorsement of (2) is, so far as we can see, stipulated rather than argued for. But step (3) in his argument is also problematic. Coyne associates the having of alternative possibilities with libertarianism. Given the traditional understanding of libertarianism, it is not surprising that he makes this assumption. For if libertarianism is true, then whenever a person has free will with respect to doing some

14 Parts of this section are taken from Timpe (forthcoming).

15 As Paul Manata has pointed out in conversation, the "if" here should clearly be an "only if." The ability to do otherwise may be necessary for free will, but clearly it's not sufficient.

particular action, it is causally open to her both to do it and not to do it. And while this is perhaps the most natural way of understanding the alternative possibilities condition, there are a number of *compatibilists* who also think that free will is primarily a function of having the ability to do otherwise. Consider, for example, Kadri Vihvelin's book, *Causes, Laws, and Free Will: Why Determinism Doesn't Matter* (2013), which defends a compatibilist account of free will that includes the ability to do otherwise. Coyne calls compatibilism a "cop-out" (2011). But most contemporary philosophers are both compatibilists and naturalists. In light of that, one might expect an *argument* against compatibilism; but it should at least make the blithe dismissal of compatibilism by Coyne and others perplexing.[16]

Coyne also thinks that free will requires substance dualism – the view, roughly, that human persons are immaterial souls or minds distinct from their physical bodies.[17] Consider the following passage:

> To assert that we can freely choose among alternatives is to claim, then, that we can somehow step outside the physical structure of our brain and change its working. That is impossible. Like the output of a programmed computer, only one choice is ever physically possible: the one you made. As such, the burden of proof rests on those who argue that we can make alternative choices, for that's a claim that our brains, unique among all forms of matter, are exempt from the laws of physics by a spooky, nonphysical "will" that can redirect our own molecules. (2012a)

The substance dualist thinks that a person's immaterial soul can causally interact with her physical body. (The difficulty in explaining how this interaction takes places is one of the main objections to substance dualism, one of which Descartes – the poster-child for substance dualism – was well aware.) If belief in free will required belief in an immaterial soul (or other supernatural "panicky metaphysics"), then it would not be surprising if naturalistic scientists like Coyne were to reject its existence. But, as Al Mele (2014) has pointed out, scientific objections to free will that are based on the claim that it requires substance dualism are misguided. As Manual Vargas puts it, "all serious accounts of free will make no appeal to substance dualism. Moreover, when one considers the empirical evidence about folk beliefs concerning the requirements of free will, the evidence seems to show that on ordinary usage, 'free will' does not require substance dualism" (2014, 217). Again, given that so few philosophers are substance dualists[18] – for example, neither of us is – this claim by Coyne and others is likewise perplexing. As we shall make clear in the final section of this paper, libertarianism does not require a commitment to substance dualism, nor to any obviously problematic non-naturalist ontology.

Consider, next, cognitive neuroscientist, best-selling author, and naturalist Sam Harris. Like Coyne, Harris thinks that free will is simply an illusion, but his reason for thinking so differs from Coyne's:

> Free will *is* an illusion. Our wills are simply not of our own making. Thoughts and intentions emerge from background causes of which we are unaware and over which we exert no conscious control. We do not have the freedom we think we have. Free will is actually more than an illusion (or less), in that it cannot be made conceptually coherent. Either our wills are determined

16 For more on Coyne's dismissal of compatibilism, see Bok (2012).

17 Substance dualism is understood as a stronger claim than just the claim that we have a soul or mind; it is the claim that we are identical with a person who is a soul or mind.

18 The most noteworthy counterexample is Swinburne (2013).

> by prior causes and we are not responsible for them, or they are the product of chance and we are not responsible for them. (Harris 2012, 5)[19]

Commenting on Harris's book, Eddy Nahmias describes the central thrust as follows: "Given his other books, one would expect science to drive Harris's conclusions, but here his argument is conceptual. Step 1: Define free will in such a way that it is impossible. Step 2: Remind us that we cannot have what is impossible" (2012, 110). But this way of approaching the matter, Nahmias writes, is naive: "My response is just as simple: Harris's definition of free will is mistaken. To have free will, people don't need the impossible; nor do most people think free will requires the impossible" (2012, 110).[20]

We think that Nahmias is correct here. Whereas Coyne's argument for the illusory nature of free will is built on the problematic assumption of determinism, Harris's argument is a problematic dilemma: either determinism is true or it's false, but in either case we lack free will. Let us explore each case in more detail.

We begin with the first horn of the dilemma: if determinism is true, then we lack free will. Harris's reasons for thinking this are much the same as Coyne's: namely, that free will requires "that each of us could have behaved differently than we did in the past" (Harris 2012, 6). That is, like Coyne, Harris identifies free will with the ability to do otherwise, which, he claims, we would lack if determinism were true. Unlike Coyne, however, Harris at least considers compatibilism:

> Today, the only philosophically respectable way to endorse free will is to be a compatibilist – because we know that determinism, in every sense relevant to human behavior, is true. Unconscious neural events determine our thoughts and actions – and are themselves determined by prior causes of which we are subjectively unaware. However, the "free will" that compatibilists defend is not the free will that most people feel they have…What does it mean to say that rapists and murderers commit their crimes of their own free will? If this statement means anything, it must be that they could have behaved differently [that is, that they could have done otherwise]. (2012, 16ff.)

But Harris accuses the compatibilist of "changing the subject" (2012, 15) and being "deliberately obtuse" (2012, 18).

There are two things worth noting here. First, a number of philosophers working in a recent movement called "experimental philosophy," sometimes in collaboration with social scientists, have begun to study ordinary folks' views about the relationship between free will and determinism (see Nahmias et al. 2006; Nichols 2006). Some of their data suggest that, contrary to what Harris claims, many – perhaps even most – of the "folk" are compatibilists. Still, the second thing to note is that even if most people were incompatibilists about free will and determinism, that wouldn't by itself mean the ordinary people were right. Establishing

19 Showing his rhetorical prowess, he calls free will not only an illusion, but also a delusion.

20 Later in the same article, Nahmias makes a similar point: "It is ironic for scientists like Harris to think that science will explain away free will, rather than helping to explain how it works. Organic chemistry did not make life disappear by explaining how living processes work. Copernicus did not explain away the earth by explaining how it moves. The only reason to think free will can't be explained is to define it such that it must be inexplicable – to assume that people demand the impossible" (2012, 113). For another critical review of Harris's book by another leading naturalist, see Dennett (2014). Dennett calls the book "valuable, as I will show, as a veritable museum of mistakes, none of them new and all of them seductive." More recently, Dennett (2013) has indicated a personal willingness to cede use of "free will" to incompatibilists.

that requires a philosophical argument. And giving such arguments is neither "changing the subject" nor being "deliberately obtuse." While we think that compatibilism is false, one needs to provide an argument for that – not merely offer a claim about what most people think.

We turn then toward the second horn of Harris's dilemma: if determinism is false, then we lack free will. What would the falsity of determinism imply for free will? Harris thinks that indeterminism is no more conducive to freedom than is determinism. If an event were not determined by previous events and the laws of nature, then such an event would be, he claims, "lucky." If an event (or action) is lucky for an individual, then she lacks control over it. And if an agent lacks control over whether or not she does some event (or action), then she does not do it freely. "The role of luck," Harris writes, "therefore, appears decisive" (2012, 4). But his treatment of the relevant philosophical literature is superficial, and he doesn't engage it in a way that justifies the inference here. There are, of course, philosophically sophisticated versions of the same kind of argument that Harris is making here. The best is likely Neil Levy's *Hard Luck* (2011), which is a wonderful treatment of the threat of luck to freedom and responsibility. What Levy's treatment illustrates, and what Harris's lacks, is an understanding of and engagement with the relevant philosophical material that doesn't make such questionable assumptions of the sort previously mentioned. But even Levy's treatment doesn't prove that indeterminism undermines free will. And Harris certainly hasn't given us any reason to think that it does.

There are, of course, naturalists who give better arguments regarding the nature of free will than do Coyne and Harris. (A great example of one such naturalist is Levy; see Chapter 22.) But the most public faces of naturalism, including Harris and Coyne, are not those who do the best work on free will, and nothing of import follows from the fact that some prominent naturalists think we lack free will. In a worthwhile article exploring the relationship between contemporary neuroscience and the traditional debates about free will, Adina Roskies writes that "current [neuroscience] knowledge has little direct bearing on this age-old philosophical problem [of free will]" (2013, 33).[21] Though a naturalist herself, Roskies too thinks that Coyne's and Harris's work on free will is rash and badly informed. This does not mean that neuroscience has no implications for how we understand free will. Roskies herself adds, for example, that "so far, the greatest effect of neuroscience has been to challenge traditional views of the relationship between consciousness and action…Converging evidence from neuroscience and psychology makes it clear that factors in addition to consciousness influence our choices" (2013, 51). While we agree that these are important findings,[22] the most plausible forms of libertarianism don't require consciousness to be the only, or even the prime, driving force behind all of our choices. A more general claim is also likely true: nothing in the recent neuroscientific findings rules out libertarianism.

21 Consider also the following: "In our volume we start from the assumption that the cognitive sciences will not be able to help us directly with the discussion between libertarians and compatibilists. However, a simple dismissal of the scientific findings as irrelevant to the free will debate would be, to say the least, premature" (Vierkant, Kiverstein, and Clark 2013, 3). Vierkant and colleagues call this "the zombie challenge": "the zombie challenge is based on an amazing wealth of findings in recent cognitive science that demonstrate the surprising ways in which our everyday behavior is controlled by automatic processes that unfold in the complete absence of consciousness…The zombie challenge suggests that the conscious self takes a backseat when it comes to the control of much of our behavior" (2013, 5). For a response to the zombie challenge, see Mele (2009). We do not think that the zombie challenge entails anything about the compatibilism/incompatibilism debate, nor about the existence of free will.

22 However, see Mele (2009) for some worries that the findings don't pose the challenge that many claim they do.

A Naturalist-Friendly Libertarian Account of Free Will

It is one thing to undermine arguments for thinking that free will is incompatible with naturalism; it is another to show that the two are compatible. In this section, we aim to do the latter by providing a consistent and coherent model for one kind of libertarian view that is naturalist-friendly.[23] We think that, in general, evaluation of metaphysical theories does not work piecemeal; it proceeds by consideration of the intuitive force and theoretical power of the more systematic, general metaphysical view in which a particular theory is embedded.[24] The same is true, we think, regarding the libertarian view we outline here. There are many aspects of the view that are contentious, and we cannot defend them adequately in the present context. We do, however, think that the larger set of metaphysical views in which the proposed account of free will is imbedded can be defended, and we will point the reader to such defenses in the notes.

We proceed as follows. First, we give reasons for thinking that all causation is substance causation. Second, we differentiate agent causation as the subspecies of substance causation that is pertinent to questions of free and responsible agency. Third, we show how the proposed framework for a libertarian view contains nothing that a naturalist as such couldn't endorse.

Substance Causation

Most contemporary philosophers writing on free will think that when an agent causes something to happen (such as taking a drink of water), the fundamental metaphysical story does not involve the agent *simpliciter*, but rather events involving the agent. It is the agent's having of certain reasons and desires, say, that cause her to choose to drink the water, not the agent as such. On this view, all causes are events. There are, however, a few philosophers who think that while some, perhaps most, causes are events, the causation involved in free and responsible agency is best understood in terms of agent causation (e.g., O'Connor 2011; Markosian 1999; 2012). In free choices, it is the agent herself, not events involving her or her parts, that is the cause of her choice. (It's this latter sort of causation, agent causation, that is purportedly "spooky," involving a "godlike" power to be an unmoved mover.) On this second view, some causes are events, but some are substances.

We think agent causation should be understood in the context of an account of the nature of causation in general. Our proposed libertarian, naturalist account of free and responsible agency rejects the two pictures of causation just discussed, and instead begins with a commitment to the metaphysical primacy of substance causation *in general*.[25] Agents are the causes of their most basic actions, but so too are electrons. While this is certainly a minority view in contemporary metaphysics, a number of metaphysicians hold that substance causation is the primary kind of causation and that purported cases of event causation can be understood as, more fundamentally, cases of substance causation.

23 For another, similar naturalistic libertarian position, see Steward (2012).

24 See Jacobs (2011). Timothy O'Connor and Jacobs make the point this way: "There is a more general moral that we hope will become more widely embraced in action theory and philosophy of mind: metaphysics matter" (2013, 190). See also Gibb (2013, 1, 15).

25 See, for further discussion, Jacobs and O'Connor (2013). Similarly, Steward (2012, esp. ch. 8) also situations her libertarian view within a substance-causal framework. While our view differs from hers in a number of ways, we are quite sympathetic to her overall approach to rational agency.

One initial motivation for this view is that, as E.J. Lowe puts it, "our ordinary ways of talking about action support, at least superficially, the idea that agent causation is a distinct species of causation" (2002, 122). We speak freely of substances as causes: the rock shattered the window, the bomb destroyed the bridge. Lowe argues, further, that "the notion of agent causation is at least *conceptually prior* to that of event causation" (2002, 134). Both Lowe and Helen Steward appeal to linguistic facts about the relation between transitive and intransitive uses of causal terms to support the legitimacy of talking about substances as causes. The transitive use of "move," as in "Sue moved her hand," implies the intransitive use in "her hand moved." And since "move" is a causal term, it follows that Sue caused her hand to be moved. According to Steward, it is in the use of these basic transitive verbs – push, pull, move, and so on – that "we really find the heart of our concept of causality" (2012, 202). In our ordinary usage of these sorts of verbs, "we find clear evidence of a commonsense metaphysics that accords to substances…a capacity to wreak effects in the world," so that "grammar reveals them to be amongst the causes *par excellence* of ordinary language" (2012, 202).

We don't take these sorts of linguistic or ordinary language arguments to establish a metaphysics of substance causation, but they should give pause to anyone who thinks that the default metaphysics of causation should be one of event causation. And they seem to establish that event causation is not the common-sense view.

Lowe also offers a more metaphysical argument for substance causation. Care needs to be taken here, as in some places Lowe takes "agent causation" to be synonymous with substance causation, while in others, as we shall, he takes more care to differentiate agent causation from the larger genus of substance causation. As an illustration of the former tendency, consider the following:

> An "agent," in the sense intended here, is a persisting object (or "substance") possessing various properties, including, most importantly, certain causal powers and liabilities. A paradigmatic example of an agent would be a human being or other conscious creature, capable of performing intentional actions. Indeed, some philosophers of action would like to restrict the term "agent" to entities such as these and, correspondingly, restrict the term "agent causation" to cases of intentional action. (Lowe 2001, 1ff.; see also Griffith unpublished)

While there is this broad sense in which all substances are agents, it is better to restrict agent causation to that subspecies of substance causation that requires reasons to play a particular role in enhancing the control of the agent in question. (More on this in the next subsection.)

Returning to Lowe's more metaphysical argument, the heart of his reasoning is this: causation is the exercise or manifestation of a power. The cause is the thing that has the power. But only substances have powers. Therefore, only substances are causes.

Lowe doesn't deny the existence of events; he just thinks that event causation can be reduced to substance causation, the latter of which is ontologically fundamental:

> I am not…denying the existence of events. Events occur when substances act or are acted upon in certain ways: indeed, events just consist in the doings of substances…The source of all change in the world lies in substances, in virtue of their causal powers and liabilities, which they characteristically exercise or manifest when they enter into suitable relationship with one another…Events as such are utterly powerless. They are mere changes in things and not the source of those changes. (2009, 342)

Lowe thinks that "basic actions" prove particularly insightful. In some cases, agents act, but not by any other means; the agent just does it.[26] In such cases, "there appears to be no suitable event involving the agent which can be called upon to provide the cause – in the event causation sense of "cause" – of the effect which, in performing that action, the agent is said to cause…Moreover, it would appear that there must be basic actions, if agents perform any action at all" (2002, 200). Granted, even if this argument is convincing, it doesn't establish that all event causation can be reduced to agent causation, only that some of it can be. But Lowe does think that this reduction, at least in terms of the causal relationship involved in purported free actions, has a payoff in terms of saving free will from the "disappearance of the agent":

> If all human agency is ultimately just a matter of one event's causing another, then, since the causal history of the events supposedly involved in any instance of human agency will plausibly be traceable back, through prior events, to times before the agent's birth, we seem to lose all sense of the agent's being genuinely responsible for – the author of – his or her own actions. A human being must then be seen as no freer, in reality, than the boulder which rolls into the tree, its rolling being caused by the action of some other object upon it, which action is in turn caused by yet earlier events – and so on back to the dawn of time. Acknowledging that there is room for a certain amount of probabilistic causation between events provides no escape from this conclusion, since a boulder's behavior would be no freer on that account than it would be in a perfectly deterministic universe. (2002, 201ff.)[27]

There is much to be said for thinking that all causation is substance causation. That view will be most compelling to those sympathetic to neo-Aristotelian approaches to metaphysics. While most naturalists work within a roughly neo-Humean approach to metaphysics, we think that there is nothing inherently non-naturalist about neo-Arisotelian approaches; indeed, often the impetus for anti-neo-Humean views of this sort comes from the field of philosophy of science (e.g., Cartwright 1989; Bird 2007). Moreover, there are serious problems with the kind of causal reductionism that is typical of neo-Humean approaches. Whereas, for neo-Humean views, causation is mere counterfactual covariance, on neo-Aristotelian views it is the exercise of causal powers by substances. Neo-Aristotelian approaches can't be ruled out from the start; indeed, we also think there are good reasons for preferring neo-Aristotelian approaches in general (and reasons for preferring neo-Aristotelian approaches to agency in particular; more on this later).[28]

So, to summarize, on the view we have in mind, substances have powers to bring about effects, and when they exercise those powers, the substances themselves are the causes of the effects.

The Role of Reasons in Agent Causation

What we've set out so far is contentious and somewhat speculative. But we don't think that an account of free will can proceed in isolation from larger metaphysical issues. We now turn more directly to free will. As mentioned before, and as is common practice, we

26 Lowe thinks an agent's waving her hand is such a basic action, but we think it more plausible that an agent's forming an intention to wave her hand is a basic action that we do not do by any other means.

27 See also Lowe (2009, 338ff.). For a worthwhile reply to the problem of the disappearing agent by an event causalist, see Franklin (2013).

28 For arguments against neo-Humean approaches, see Jacobs and O'Connor (2013, esp. §§ 1, 2). For a defense of one particular neo-Aristotelian approach, see Jacobs (2010; 2011).

understand free will as the control condition on moral responsibility. On our view, free will is a system-level feature, for reasons that should become clear momentarily, that involves the faculties of the will and the intellect.[29] The will is a kind of causal power that enables a substance (or agent) that has it to cause certain types of event, namely actions. Actions are a distinctive kind of event, and agent causation is a distinctive kind of substance causation. We agree with Lowe that what differentiates agent causation from the larger species of causation is that agent causation is inherently "executive" or "volitional," and that this executive capacity of the agent is intimately connected with the agent's responsiveness to reasons.[30]

Here, we begin to see differences between our view and Lowe's. According to Lowe, "What, in my view, distinguishes the will from any other kind of spontaneous power is (1) that it is a two-way power – a power either to will or not will a particular course of action – and (2) that it can be exercised rationally, that is, 'in the light of reason'" (2013, 154; see also Griffith unpublished). We differ from Lowe in a number of important ways. First, we do not think that the exercise of agency requires, at the time of each act of agency, one to have the ability to will or refrain from willing that particular course.[31] This difference isn't crucial for our present purposes. What is particularly important here, and what differentiates the will from other causal powers, is the will's responsiveness to reasons, which are provided by the intellect.[32]

More importantly, we disagree about the role of reasons in free action. We think, like Lowe, that reasons play an ineliminable role in free agency, though we think they play a different role than does Lowe. We also differ from Lowe about the exact nature of the relationship of reasons to our agency. Lowe is an externalist about reasons. While we are inclined to reject reasons externalism, nothing in this chapter is incompatible with it. Our view here does require the view that reasons are not causes; that is, we reject the view that when you act it's your reasons, or the having of them, that are the causes of your actions. On our view, reasons are relevant, but not in a directly causal way. So what does it mean to be reasons-responsive, to act for a reason? On our view, reasons dispose agents to act in certain ways.[33] When an agent acts on the basis of a reason, the agent is manifesting her agential dispositions. If she had had different reasons, she would have been differently disposed. But these reasons, as well as the agent's ability to detect them and choose on the basis of them, are features of the agent. And if, as we think, agents are natural substances, then their features are natural too: they are features of biological organisms.[34]

29 "Faculties" need not be mysterious, non-naturalistic entities; they are simply powers or capacities of the agent.

30 See Lowe (2009, 347ff.).

31 For more on why this is, see Timpe (2013, esp. ch. 6).

32 For an account of the role that reasons – in particular, reasons related to perceived goodness – play in the exercise of free will, see Timpe (2013, ch. 2).

33 There are further related issues in the vicinity, such as the relationship between acting on a reason and acting for a reason. We are avoiding these issues both due to space constraints and because we don't think that their resolution matters directly for naturalism. See O'Connor (2008) for further discussion.

34 If these features are features of parts of an agent, rather than of the agent *simpliciter*, then we may face the problem of the disappearing agent once again, since it's not clear that she is the cause, instead of being the place where the causing occurs. For this reason, we may need to think of some of the agential capacities in holistic terms, or in terms of "emergence." We see no reason, again, to think that if a large object has properties that are not made up of properties of the object's part, then naturalism is false. For an account of emergent properties and the substances that have them, see O'Connor and Jacobs (2003) and O'Connor and Wong (2005).

Naturalism

In Chapter 1, Kelly James Clark lays out a number of forms of naturalism. He describes "metaphysical naturalism" as the ontological thesis that no non-supernatural entities exist. While, historically, agent causation has been thought to require a non-natural entity such as an immaterial soul (thus, also requiring a different kind of causation than is found in the natural order), the previous sections should show that an immaterial soul is not required. If all causation is substance causation, then agent causation (a subspecies of substance causation) need not be problematic for the naturalist. If everything is spooky, nothing is. And if all causation is ultimately substance causation, then it's false that science doesn't appeal to substance causation.

Furthermore, reasons-responsiveness is not ruled out by what, in Chapter 1, Clark calls "broad naturalism." Granted, there are naturalistic positions that would be incompatible with the view that we've offered – if one rejected the existence of reasons or moral responsibility, for example, or perhaps scientific naturalism. But these views go beyond mere naturalism, and we see no reason to endorse their further claims.

Conclusion

We've offered an agent-causal libertarian view that is consistent with naturalism. We've also offered reasons in support of the various elements of that view. We have, for example, given reasons why one might think that all causation is substance causation and that neither substance causation in general nor agent causation in particular is ruled out by naturalism. This, of course, doesn't mean that agent causation is true. But we think it does show that a commitment to naturalism by itself doesn't rule out the existence of free will, even if free will is of the agent-causal sort that we've suggested.

Though both of us are incompatibilists (i.e., we think that the truth of determinism would entail that we lack the kind of freedom at issue in the philosophical debates about moral responsibility), this is independent of our argument here. Moreover, the debate between incompatibilists and compatibilists is a philosophical argument, and not one that can be settled by science. However, we think that it is worth saying something further regarding libertarian views of free will in comparison to their compatibilist competitors.

The most compelling philosophical objection to libertarian views of free will is what has come to be called "luck objection." There is no single objection that goes by this moniker; rather, the luck objection is best thought of as a family of related objections. But here is one representative expression:

> Another frequently heard objection to indeterministic free will is precisely that undetermined free choices must *always* amount to mere *random* choices, like flipping a coin or spinning a wheel to select from among a set of alternatives. Perhaps there is a role for random choices in our lives – for sometimes settling choices by a coin flip or spinning a wheel – when we are indifferent to the outcomes. (Which movie should I see tonight when I like both available options?) But suppose that *all* our free and responsible choices – including momentous ones, like whether to act heroically or treacherously, to lie to a friend, or to marry one person rather than another – had to be settled by random selection in this way. Such a consequence, according to most philosophers, would be a reduction to absurdity of the view that free will and responsibility require indeterminism. (Kane 2005, 37)

In our view, the luck objection is a significant but not unanswerable challenge, even though we don't have the space to go into how we think it should be addressed here. But, in brief, we don't think that there is any good reason to think that all cases of mere indeterminism are the kind of control-undermining luck that would undermine free will.[35]

What's more important for our present concerns is the importance of the compatibility of free will with indeterminism. Even if the presence of the indeterminism that libertarians think is required doesn't undermine free will, as the luck objection holds, many suggest that this requirement is a disadvantage for the libertarian. John Fischer, for example, writes that compatibilism's impunity at this point is a significant point in its favor. Thus, compared with incompatibilism, compatibilism is able to reconcile freedom with a greater range of scientific discoveries. But is this so? It depends on the kind of compatibilism under discussion.

A number of compatibilists think that free will not only is compatible with determinism but actually requires it. This kind of compatibilism is likewise hostage to potential threats by scientific discovery, as is libertarianism insofar as it too requires a particular view regarding the truth or falsity of determinism. In contrast, most compatibilists are, as mentioned earlier, "supercompatibilists": they think that free will neither requires nor is ruled out by determinism. Supercompatibilism is less threatened by any potential discoveries of science regarding determinism.

But note the following. The more the compatibilist thinks that free will is compatible with both determinism and indeterminism, the less motivated will be the luck objection to libertarianism, since at the heart of that objection is the belief that indeterminism undermines rather than supports freedom by making the agent's choices "simply the result of luck." If one thinks that free will is compatible with both determinism and indeterminism at a wide range of places causally relevant to a particular action, the luck objection will carry less dialectical force – for if indeterminism *per se* doesn't undermine free will, then the driving intuition behind the luck objection will be undercut.[36]

Suppose that we're wrong in that free will requires the falsity of determinism. If so, then the view of the nature of free will as centrally involving a reasons-responsive form of agent causation given here will not be undermined: it can simply be understood as compatible with determinism. And there are already compatibilist versions of agent causation.[37]

In our view, contrary to what a number of naturalists have claimed, the truth of naturalism wouldn't rule out free will, not even of the agent-causal kind. Whether or not agent-causal free will is ultimately of the libertarian or compatibilist variety will depend on the outcome of philosophical debate: it's not something settled by the debates over naturalism. Like Levy, we think that the "prospects for a fully naturalistic account of free will are therefore tied up with the prospects for fully naturalistic accounts of other properties" (Chapter 22). But we see no reason to think this cannot be accomplished.

Acknowledgments

Thanks to Kelly James Clark and Meghan Griffith for helpful comments on an earlier version of this paper.

35 For an argument that along these lines, see Timpe (2012, ch. 10).

36 For more on this issue, see Timpe (2010).

37 See Markosian (1999; 2012). We see it as a benefit of our view that the substance-causal metaphysic of free will we outline here is available to both the compatibilist and the incompatibilist.

References

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Bok, H. (2012). Want to Understand Free Will? Don't Look to Neuroscience. *Chronicle of Higher Education*. Available from: http://chronicle.com/article/Hilary-Bok/131168/ (last accessed July 15, 2015).

Cartwright, N. (1989). *Nature's Capacities and their Measurements*. Oxford: Oxford University Press.

Clarke, R. (2002). Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency. In *The Oxford Handbook of Free Will,* edited by Robert Kane, pp. 356–385. Oxford: Oxford University Press.

Coyne, J. (2011). More on Free Will: Dr. Pigliucci Weighs In, I Respond. *Why Evolution is True*. Available from: http://whyevolutionistrue.wordpress.com/2011/12/04/more-on-free-will-dr-3-pigliucci-weighs-in-i-respond/ (last accessed July 15, 2015).

Coyne, J. (2012a). Why You Don't Really Have Free Will. *USA Today*. Available from: http://usatoday30.usatoday.com/news/opinion/forum/story/2012-01-01/free-will-science-religion/52317624/1 (last accessed July 15, 2015).

Coyne, J. (2012b). You Don't Have Free Will. *Chronicle of Higher Education*. Available from: http://chronicle.com/article/Jerry-A-Coyne/131165/ (last accessed July 15, 2015).

Dennett, D.C. (2013). A Dialogue on Free Will. *Methode* 2.3: 46–68.

Dennett, D. (2014). *Reflections on Free Will*. Available from: http://www.naturalism.org/Dennett_reflections_on_Harris's_Free_Will.pdf (last accessed July 15, 2015).

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Fischer, J.M. (1999). Recent Work on Moral Responsibility. *Ethics* 110: 93–139.

Fischer, J.M. (2000). Problems with Actual-Sequence Incompatibilism. *Journal of Ethics* 4: 323–328.

Fischer, J.M. (2006). The Cards that are Dealt You. *Journal of Ethics* 10: 107–129.

Fischer, J.M. (2007). Compatibilism. In *Four Views on Free Will*, edited by Derk Pereboom, John Martin Fischer, Robert Kane, and Manuel Vargas, pp. 44–84. Malden: Blackwell.

Fischer, J.M. and Ravizza, M. (1998). *Responsibility and Control*. Cambridge: Cambridge University Press.

Franklin, C.E. (2013). Event-Causal Libertarianism, Functional Reductionism, and the Disappearing Agent Argument. *Philosophical Studies* 170: 431–432.

Gibb, S.C. (2013). Introduction. In *Mental Causation and Ontology*, edited by E.J. Lowe, S.C. Gibb, and R.D. Ingthorsson, pp. 1–17. Oxford: Oxford University Press.

Griffith, M. (Unpublished). Substance Causation.

Harre, R. (2001). Active Powers and Powerful Actors. *Philosophy: The Journal of the British Institute of Philosophical Studies* 48(Suppl.): 91–109.

Harris, S. (2012). *Free Will*. New York: Free Press.

Holton, R. (2013). From Determinism to Resignation; and How to Stop It. In *Decomposing the Will*, edited by Andy Clark, Julian Kiverstein, and Tillmann Vierkant, pp. 87–100. Oxford: Oxford University Press.

Howard, G. (2008). Whose Will? How Free? In *Are We Free? Psychology and Free Will*, edited by Roy F. Baumeister, John Baer, and James C. Kaufman, pp. 260–273. Oxford: Oxford University Press.

Jacobs, J.D. (2010). A Powers Theory of Modality: Or, How I Learned to Stop Worrying and Reject Possible Words. *Philosophical Studies* 151(2): 227–248.

Jacobs, J.D. (2011). Powerful Qualities, Not Pure Powers. *Monist* 94(1): 81–102.

Jacobs, J.D. and O'Connor, T. (2013). Agent Causation in a Neo-Aristotelian Metaphysics. In *Mental Causation and Ontology*, edited by E.J. Lowe, S.C. Gibb, and R.D. Ingthorsson, pp. 173–192. Oxford: Oxford University Press.

Kane, R. (2001). Introduction: The Contours of the Contemporary Free Will Debates. In *The Oxford Handbook of Free Will*, edited by Robert Kane, pp. 3–42. Oxford: Oxford University Press.

Kane, R. (2005). *A Contemporary Introduction to Free Will*. New York: Oxford University Press.
Levy, N. (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.
Lowe, E.J. (2000). *An Introduction to the Philosophy of Mind*. Cambridge: Cambridge University Press.
Lowe, E.J. (2001). Event Causation and Agent Causation. *Grazer Philosophische Studien* 61: 1–20.
Lowe, E.J. (2002). *A Survey of Metaphysics*. Oxford: Oxford University Press.
Lowe, E.J. (2009). Free Agency, Causation, and Action Explanation. In *New Essays on the Explanation of Action*, edited by Constantine Sandis, pp. 338–355. London: Palgrave MacMillan.
Lowe, E.J. (2013). Substance Causation, Powers, and Human Agency. In *Mental Causation and Ontology*, edited by E.J. Lowe, S.C. Gibb, and R.D. Ingthorsson, pp. 153–172. Oxford: Oxford University Press.
Markosian, N. (1999). A Compatibilist Version of the Theory of Agent Causation. *Pacific Philosophical Quarterly* 80: 257–277.
Markosian, N. (2012). Agent Causation as the Solution to All the Compatibilist's Problems. *Philosophical Studies* 157: 383–398.
McKenna, M. (2008). Putting the Lie on the Control Condition for Moral Responsibility. *Philosophical Studies* 139: 29–37.
Mele, A.R. (2009). *Effective Intentions: The Power of Conscious Will*. Oxford: Oxford University Press.
Mele, A. (2014). Free Will and Substance Dualism: The Real Scientific Threat to Free Will? In *Moral Psychology*, edited by Walter Sinnott-Armstrong, vol. 4, pp. 195–208. Cambridge: Cambridge University Press.
Nahmias, E. (2012). Do We Want the Impossible? Review of Sam Harris' *Free Will*. *The Philosophers' Magazine*.
Nahmias, E. (2014). Is Free Will an Illusion? Confronting Challenges from the Modern Mind Sciences. In *Moral Psychology*, edited by Walter Sinnott-Armstrong, vol. 4, pp. 1–26. Cambridge: Cambridge University Press.
Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2006). Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research* 73: 28–53.
Nichols, S. (2006). Folk Intuitions on Free Will. *Journal of Cognition and Culture* 6(1/2): 57–86.
O'Connor, T. (2008). Agent-Causal Power. In *Dispositions and Causes*, edited by Toby Handfield. Oxford: Oxford University Press.
O'Connor, T. (2011). Agent Causal Theories. In *The Oxford Handbook of Free Will*, 2nd edn., edited by Robert Kane. Oxford: Oxford University Press.
O'Connor, T. and Jacobs, J.D. (2003). Emergent Individuals. *The Philosophical Quarterly* 53(213): 540–555.
O'Connor, T. and Wong, H.Y. (2005). The Metaphysics of Emergence. *Noûs* 39(4): 658–678.
Papineau, D. (2001). The Rise of Physicalism. In *Physicalism and its Discontents*, edited by Carl Gillett and Barry Loewer, pp. 3–36. Cambridge: Cambridge University Press.
Papineau, D. (2009). The Causal Closure of the Physical and Naturalism. In *The Oxford Handbook of Philosophy of Mind*, edited by Ansgar Beckermann, Brian P. McLaughlin, and Sven Walter, pp. 53–65. Oxford: Oxford University Press.
Roskies, A.L. (2013). The Neuroscience of Volition. In *Decomposing the Will*, edited by Andy Clark, Julian Kiverstein, and Tillmann Vierkant, pp. 33–59. Oxford: Oxford University Press.
Steward, H. (2012). *A Metaphysics for Freedom*. Oxford: Oxford University Press.
Swinburne, R. (2013). Mind, Brain, and Free Will. Oxford: Oxford University Press.
Timpe, K. (2010). Demotivating Semicompatibilism. *Ideas y Valores: Revista colombiana de fiosofia* 58: 109–124.
Timpe, K. (2012). *Free Will: Sourcehood and Its Alternatives*, 2nd edn. London: Bloomsbury.
Timpe, K. (2013). *Free Will in Philosophical Theology*. London: Bloomsbury.
Timpe, K. (Forthcoming). Agency and Neuroscience. The Table Journal, Biola University Center for Christian Thought.

van Inwagen, P. (2008). How to Think about the Problem of Free Will. *Journal of Ethics* 12: 327–341.

Vargas, M. (2007a). Response to Kane, Fischer, and Pereboom. In *Four Views on Free Will*, edited by Derk Pereboom, John Martin Fischer, Robert Kane, and Manuel Vargas. Oxford: Blackwell.

Vargas, M. (2007b). Revisionism. In *Four Views on Free Will*, edited by Derk Pereboom, John Martin Fischer, Robert Kane, and Manuel Vargas, pp. 126–165. Oxford: Blackwell.

Vargas, M. (2012). Why the Luck Problem Isn't. *Philosophical Issues* 22.1: 419–436.

Vargas, M. (2013). *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press.

Vargas, M. (2014). Reconsidering Scientific Threats to Free Will. In *Moral Psychology*, edited by Walter Sinnott-Armstrong, vol. 4, pp. 217–226. Cambridge: Cambridge University Press.

Vierkant, T., Kiverstein, J. and Clark, A. (2013). Decomposing the Will: Meeting the Zombie Challenge. In *Decomposing the Will*, edited by Andy Clark, Julian Kiverstein, and Tillmann Vierkant, pp. 1–30. Oxford: Oxford University Press.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford: Oxford University Press.

Zagzebski, L. (1990). What If the Impossible had been Actual? In *Christian Theism and the Problems of Philosophy*, edited by Michael D. Beaty, pp. 165–183. Notre Dame: University of Notre Dame Press.

24

Does the New Wave in Moral Psychology Sink Kant?

VALERIE TIBERIUS

The longstanding debate between Humeans and Kantians has been fueled by recent work in moral psychology that draws on empirical findings in psychology, neuroscience, and other fields. The empirical turn has not been kind to Kant. The psychologist Jonathan Haidt calls Kant's worship of reason a "rationalist delusion" (2012, 34), and one can easily get the impression that the Kantian project in ethics is as dead as the man who inspired it. Is this true? Has the new wave in moral psychology caused the Kantian boat to capsize? In this chapter, I distinguish some different Kantian claims that are under fire and I assess the evidence against them. Though I am not a Kantian, I argue that rumors of the death of Kantianism have been greatly exaggerated.

Metaethics: The Empirical Critique of Rationalism

The Kantian position is complex and nuanced, and there is no way to do it justice in a single chapter. Instead, this chapter will take for granted some familiarity with Kantianism and will assume a general definition of rationalism about moral judgment that will allow us to make sense of and evaluate the empirical attack on rationalism. Let's define rationalism about moral judgment as the view that the truth of a moral judgment is determined by rational principles. The Kantian rationalist also accepts that moral judgments give us reasons that motivate us insofar as we are rational (independently of our nonrational sentiments or desires). What does the empirical evidence tell us about this picture?

Emotions Influence and Cause Moral Judgments

We can start with evidence for sentimentalism, which is often taken to be evidence against rationalism. According to sentimentalism, emotions play an essential role in moral judgment. Sentimentalists differ over whether moral judgments are reports about or expressions of our sentiments, and over which sentiments under what conditions are crucial to

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

moral judgment. Despite these disagreements, sentimentalists all agree that moral judgments are not made true by rational principles. For example, if moral judgments are reports about our emotions, then they are made true by facts about our emotions (perhaps facts about the emotional response we would have in certain conditions). Or, if moral judgments are *expressions* of our emotions, then whatever story might be told about whether and in what sense moral judgments are truth apt, it couldn't be a story that appeals to rational principles (Blackburn 1984; Gibbard 1992). Sentimentalism of either form presents a picture of moral judgment that is at odds with the rationalist picture, according to which moral judgments do sometimes (when true) track the deliverances of reason.

The case for sentimentalism, particularly the case defended by Jesse Prinz (2007), often starts with evidence that emotions influence and cause moral judgment, of which there is a good deal.[1] To give one colorful example, here's an experiment that shows the effect of disgust on moral judgment. Psychologists asked participants to answer questions about the moral propriety of four different scenarios, two having to do with incest between first cousins, one having to do with the decision to drive rather than walk to work, and the last having to do with a studio's decision to release a morally controversial film (Schnall et al. 2008). The participants were divided into three different groups: no-stink, mild-stink, and strong-stink, distinguished by the amount of stink – in the form of "commercially available fart spray" sprayed into a nearby trash can – in the environment. The results were that feelings of disgust increase people's tendency to make harsh moral judgments. Other experiments have shown that anger makes people more punitive and harsh in their moral judgments about crimes against persons (Lerner, Goldberg, and Tetlock 1998; Seidel and Prinz 2013).

There is also evidence that emotions cause us to make moral judgments that we would not otherwise make. For example, Thalia Wheatley and Jonathan Haidt hypnotized half the participants in one study to feel disgust when they heard the word "often" and the other half to feel disgust when they heard the word "take." All the participants then read some scenarios, one of which was this:

> Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He {tries to take/often picks} topics that appeal to both professors and students in order to stimulate discussion. (Wheatley and Haidt 2005)

Participants were randomly assigned to a group that got one or the other of the phrases in the square brackets; no student saw both the phrases in the brackets.

To those of us who have not been hypnotized, it doesn't seem like Dan has done anything wrong. Moreover, participants who read the scenario that did not contain their disgust-inducing word did not rate Dan's behavior as wrong. However, for the students who did feel disgust (because they read the scenario with the word that induced disgust in them), there was a tendency to rank Dan's actions as wrong. This is a case in which the people in question would not have made a judgment of moral wrongness at all were it not for the emotion of disgust they experienced.

The fact that emotions *influence* moral judgments does not establish that moral judgments *are* emotional responses, nor even that emotions are an essential part of moral judgment. (Nor does Prinz think it does – he offers this evidence as part of a larger case). This would only be an argument for sentimentalism if the sentimentalist understanding of moral judgment were

1 But see Jones (2006) for problems with Prinz's arguments.

the *only* way to explain the influence of emotions on moral judgment. Other explanations are possible; it could be that emotions influence moral judgment in the way that wearing rose-colored glasses can influence your judgment about the color of the sky: the glasses sway your judgment, but they're not part of the content of the judgment that the sky is pink. Evidence that emotions cause moral judgments is somewhat more difficult to explain away, but the person who rejects sentimentalism can still argue that when moral judgments are entirely caused by emotions, they are akin to manipulated illusions; after all, it has not been shown that *all* of our moral judgments are such that we would not make them were it not for our emotions.

Perhaps the sentimentalist would have a stronger argument if there were evidence that we simply *cannot* make moral judgments without emotions. We'll consider this possibility in the next section.

Emotions are Necessary for Moral Judgment

Some have thought that psychopaths provide evidence that we cannot make moral judgments without emotions, because (to oversimplify) psychopaths are amoral and they do not experience normal emotions like sympathy or compassion.[2] If we cannot make moral judgments without emotion, it might seem like emotions are essential to moral judgment in a way that supports the sentimentalist characterization.

Psychopathy is a personality disorder characterized by lack of empathy, impulsivity, egocentrism, and other traits. The disorder is often diagnosed by the Psychopathy Checklist, which asks a number of questions that cluster under the headings "aggressive narcissism" and "socially deviant lifestyle" (Hare and Vertommen 2003). Because psychopaths lack empathy, they are of interest to those who think emotions like empathy are essential to moral judgment. Shaun Nichols (2010), for example, thinks that the evidence from psychopathy counts against rationalism because psychopaths do not have defects of reasoning and yet do not seem to make moral judgments in the same way that the rest of us do.[3] The basic argument goes this way:

(1) Psychopaths do not make a distinction between moral wrongs and conventional wrongs.
(2) It is the defect to the emotional response system that is responsible for psychopaths' decreased ability to distinguish moral wrongs from conventional wrongs.
(3) Therefore, a functioning emotional response system is essential to moral judgment.

The conclusion of this argument is taken to be strong evidence for sentimentalism.

Let's consider the steps of this argument in detail. The first thing to notice is the importance of the distinction between "moral" and "conventional." Conventional norms, such as "You shouldn't go outside in your pajamas," are different from moral norms in a variety of ways. Moral norms are thought to be more serious and to have wider applicability than conventional norms. Conventional norms are thought to be contingent on an authority (such as a teacher or the law, or, in the case of the pajamas, a culture), and they receive a

2 For brevity, I'll talk about "the psychopath," even though this is misleading. In reality, people called psychopaths are a rather varied group who score higher or lower on different diagnostic criteria for psychopathy, and who differ in terms of their capacity to understand moral norms (Aharoni, Sinnott-Armstrong, and Kiehl 2012).

3 See Roskies (2003) for an opposing viewpoint.

different kind of justification from moral norms, which are often justified in terms of harm or fairness (Nichols 2004). For example, young children will say that it would be wrong to pull another child's hair, even if the teacher said it was o.k., because pulling hair hurts, whereas the wrongness of chewing gum in class depends on the teacher's forbidding it.

It is widely believed that psychopaths don't really understand this distinction (Blair 1995); that is, psychopaths think of what's morally wrong as what's prohibited by the local authority and do not see moral transgressions as more serious than other kinds of violation of rules. This claim is now considerably more controversial than it used to be, but even the latest research confirms that the "affective defect" part of psychopathy does predict poor performance in distinguishing moral from conventional wrongs (Aharoni, Sinnott-Armstrong, and Kiehl 2012). Because they don't feel bad when others suffer, "they cannot acquire empathetic distress, remorse, or guilt. These emotional deficits seem to be the root cause in their patterns of antisocial behavior" (Prinz 2006). Further, these emotional deficits seem to be responsible for the fact that they don't make the same kinds of moral judgments that the rest of us do.

Jeanette Kennett (2006) defends rationalism against this sentimentalist attack by arguing that psychopaths have *rational* defects that explain their inability to make real moral judgments or to be motivated by them. According to Kennett, the evidence suggests

> that psychopaths have at best a weak capacity to stand back from and evaluate their desires, to estimate the consequences of their actions, to eschew immediate rewards in favour of longer term goals, to time order, to resolve conflicts among their desires, to find constitutive solutions. To these rational shortcomings we may add that psychopaths frequently choose grossly disproportionate means to their immediate ends or fail to adopt the necessary means to their proclaimed ends. (2006, 77)

She argues that these rational defects play a crucial role in the psychopath's moral defects. Further, Kennett argues, even if the relevant defect for the psychopath is an emotional defect, this need not embarrass the rationalist. After all, if the rationalist's primary claim is about what makes moral judgments true, she can be agnostic about what mechanism is operating when we make moral judgments. In other words, the rationalist can accept that a properly tuned emotional system is necessary for making appropriate moral judgments, while maintaining that what is distinctive about moral judgment is that they are justified by rational principles.

The evidence from psychopathy, then, is not decisive against Kantianism. First of all, psychopathy is a tricky category that includes multiple defects, not all of which are emotional, and it is a complex matter to figure out what is responsible for what. Second, even if we agree that what is important is the ability to distinguish moral from conventional wrongs, and even if we agree that psychopaths with emotional defects are thereby hindered in their ability to make this distinction, we have not shown that reasoning plays no role in the making of moral judgments, nor that moral judgments are not to be assessed in terms of how well they track rational moral principles.

Reasoning is Not Necessary for Moral Judgment

We might take a further stab at defeating rationalism by looking at evidence that no reasoning is required for making moral judgments. Jonathan Haidt, famous for this attack on rationalism, contends that moral judgments cannot be rational judgments because we do not typically engage in reasoning when we make them.

Haidt (2001) argues that reasoning does not play the causal role in producing moral judgment that we once thought it did. He claims that moral judgments are typically made intuitively, on the basis of sentiments (which he calls "intuitions"). While it is possible for us to reason about our moral judgments, according to Haidt, this happens rarely. He calls his theory of moral judgment "the social intuitionist model" (SIM), because moral judgments are quick and intuitive, and when reasoning is used, it is usually social reasoning that takes place as people talk and argue with each other to try to figure things out. SIM does allow that individual reasoning or "private reflection" occurs and can affect our judgments, but it is not the typical cause of moral judgment.

We can't review all of Haidt's evidence here, but one piece has attracted a great deal of discussion by philosophers and is therefore worth examining in some detail. This is the phenomenon of dumbfounding, which states that we can (and often do) construct justifications for intuitive judgments that were not made by reasoning. This creates the illusion of objective reasoning when what is really happening is *post hoc* rationalization.

Moral dumbfounding happens when a person cannot find any reasons for the moral judgment she makes and yet continues to make it anyway. In the study that introduced the phenomenon, subjects were presented with the following scenario:

> Julie and Mark, who are brother and sister, are traveling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. So what do you think about this? Was it wrong for them to have sex? (Bjorklund, Haidt, and Murphy 2000)

Most subjects say that the siblings' behavior is wrong, and they offer reasons for their judgment. They say Mark and Julie may have a deformed child, or that it will ruin their relationship, or cause problems in their family, and so on. But because of the way the scenario is constructed, the interviewer can quickly dispel their reasons, which leads to the state of dumbfounding. According to Haidt, in an interview about his findings, dumbfounding only bothers certain people:

> For some people it's problematic. They're clearly puzzled, they're clearly reaching, and they seem a little bit flustered. But other people are in a state that Scott Murphy, the honors student who conducted the experiment, calls "comfortably dumbfounded." They say with full poise: "I don't know; I can't explain it; it's just wrong. Period." (Sommers 2005)

Dumbfounding, the argument goes, shows that most people don't make moral judgments for reasons. Rather, people offer *post hoc* rationalizations of their emotional convictions, and when these rationalizations are undermined, they stick with their convictions anyway.

Does Haidt's research on the *causes* of moral judgment present problems for Kantians? While Kantian rationalists don't make claims about the typical causes of moral judgment, perhaps they make assumptions that are undermined by Haidt's research. This is the question we will now explore.

First, let's consider whether Haidt and the Kantians mean the same thing by "reasoning." If they do not, then Haidt's challenge won't necessarily undermine rationalism. Haidt's

picture of moral reasoning seems rather different from what Kantians take moral reasoning to be. Haidt concedes rare cases in which people "reason their way to a judgment by sheer force of logic" (Haidt and Bjorklund 2008, 819). But this is a caricature: moral rationalists do not typically think that we reason ourselves into moral positions by sheer force of logic. One tool of moral reasoning that Kantians think is particularly important is universalization: when we're unsure what to do, we should ask ourselves whether the intention of our action requires making a special exception for ourselves, or whether it is an intention that we think is acceptable for everyone to have. This sort of reasoning is not the sheer force of logic.

Still, there is a good case against the idea that we often engage in slow, deliberate conscious reasoning when we make moral judgments, and Kantians do seem to think that this sort of reasoning is important. But Kantians don't claim that we routinely engage in this kind of reasoning, nor do they claim that our moral judgments are typically caused by reasoning. The Kantian holds that some moral judgments (the correct ones) are backed up by rational principles and that we *could* – if we needed to – use our rational capacities (such as universalization) to justify them. This doesn't require that we always, or even typically, reason to our moral judgments. Indeed, it would be a waste of our precious cognitive resources to do this, since most of our moral judgments are fairly easy and uncontroversial. When you read in the paper that someone has stolen billions of dollars from a retirement fund, or that someone has sold 10-year-old children into sexual slavery, you find yourself making moral judgments about these people. There's no need for reasoning here; reasoning would be wasted effort, since the cases are so obvious. Reasoning, however, is needed in cases of conflict when we aren't sure what to do. For example, what should you do if you rear end someone's car in a parking lot when no one else is looking, causing a small amount of damage? Of if you discover your very good friend cheating on his taxes? Your automatic judgment might be to do nothing (to drive away, to turn a blind eye for the sake of your friendship), but if you think about it, you might conclude that this wouldn't be right.

So, Kantians do not need to assume that our moral judgments are always caused by reasoning. What they do assume is that moral judgments can be justified and that reasoning – when it's done well – produces justification. On the Kantian view, then, while we could reason our way to a moral judgment if we needed to, it's not a problem if many of our moral judgments are fairly automatic. Haidt concedes that we sometimes arrive at judgments through private reasoning. He also thinks that we engage in social reasoning: reasoning with each other in the form of argument and gossip. Kantians do not need to assume that moral reasoning is always private. Indeed, reasoning with each other might help us overcome our biases, so that we can be more impartial and better universalizers.

We Are Not Reflective Creatures

Dumbfounding raises another problem, one concerning the Kantian assumption about the kinds of agents we human beings are. The assumption is that we are reflective creatures who aim to act for reasons that we can endorse as justifications of our actions. According to Christine Korsgaard (1996), ethics is a response to the practical problem that confronts us as reflective creatures, a problem that any normal human being will recognize: the problem of *what to do*. When confronted with a quandary about what to do, "[t]he reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward" (Korsgaard 1996, 93). When our desires or inclinations conflict with each other or with what we deem morally right, we

need a conclusive reason to go one way rather than another, and that reason can't be another desire or inclination. Rather, reasons that answer this very practical problem must be underwritten by a rational principle or law.

Do we have reflective minds like this? Do we need a reason or a consideration that provides a justification for our action? And is it impossible for this reason to depend on a desire or a sentiment?

Dumbfounding seems to afford evidence against the first claim. A "comfortably dumbfounded" person seems perfectly happy without reasons for their moral judgments. If this is correct, it contradicts not just the view of a few Kantians, but also a general and very widely held assumption about moral judgments made by most philosophers (Kennett and Fine 2008). Introductory ethics textbooks and Introduction to Ethics courses often begin by distinguishing moral judgments from mere judgments of taste by pointing out that we have justifying reasons for the former but not for the latter. "You morally ought not to eat kale" is importantly different from "Kale tastes awful," for example, because the first implies that there are some reasons that make eating kale a bad idea, not just for someone who doesn't like kale or is allergic, but for anyone. Sentimentalists and rationalists alike have agreed with this, and sophisticated sentimentalists have bent over backwards trying to accommodate the idea that we think that our moral judgments should be backed up with reasons (Blackburn 1998; Gibbard 1992). So, it's worth thinking about what dumbfounding really shows. To cut to the chase, I don't think the evidence shows that there's anything wrong with the conception of a moral judgment as involving justifying reasons (Tiberius 2013).

To doubt that people care about the reasons for their moral judgments, we would have to see that people:

(1) Do not think there are reasons for their judgments (as opposed to thinking that there are reasons, but not knowing what the reasons are).

and:

(2) Are entirely unperturbed by this fact and have no inclination to reconsider their judgments.

We do not know that (1) is true from the existing evidence. The claim "I don't know why it's wrong, it's just wrong" is ambiguous between "It's wrong for some reason that I don't know" and "There's no reason it's wrong, it just is!" It is also possible that people think some things are self-evidently wrong and that they take this self-evidence to be a reason. We also do not know that (2) is true. The fact that people are unwilling to change their judgments in an artificial interview setting does not mean that they never feel pressure to change them. It may take time for shaken confidence to cause someone to change their judgments. We would need evidence of how people respond to the challenge to provide justification over the long term and in real-life settings. Further, even if a person were *never* to change her judgment when she discovered she had no reasons for it, this would not show that she didn't care about reasons. If people don't change their judgments when nothing is at stake or when they don't feel like thinking about it at the time, this does not count against the claim that people care whether their judgments are justified when it counts.

The argument I've offered only establishes that *it hasn't yet been shown* that people don't care about reasons. But it is unlikely that studies will ever show that nobody has a conception

of a moral judgment as one that is held for justifying reasons, unlike judgments of mere taste. Reflection on moral disagreement indicates that people often think of moral judgments as different from judgments of taste. People cast their votes for political candidates who agree with them about the morality of abortion, they march in the streets to protest or support gay marriage, they donate money to charities that help stray animals or feed hungry people. It's hard to imagine that such morally committed people don't think there are reasons for the moral judgments upon which they are acting. Of course, such thoughts are not always correct, and people don't always care about justification. But the idea that moral judgments are different from judgments of taste (like "Kale tastes awful") with respect to their being supported by reasons is not just the crazy idea of a few philosophers: it is a pervasive idea in general. Indeed, many philosophers (Kantians and non-Kantians alike) would argue that those who deny that moral judgments require justification do not grasp what is distinctive about them.

We have been discussing the first of the Kantian claims mentioned earlier, namely, that the reflective mind needs a reason. The second claim is that the reason cannot be ultimately explained by a desire or a sentiment. The rationalist thinks that reasons (genuine reasons, ones that justify what we do) must be backed up by rational principles, not desires or sentiments. Their "authority" or justificatory weight cannot ultimately be grounded in personal preference, such as how we feel or what we care about. This argument seems to assume something about our psychology, namely, that we are reflective creatures of a particular kind: ones whose conviction that we have reasons to act rests fundamentally on the idea that there are universal principles that support these reasons. In other words, reflective creatures like us need to find something at the bottom of the pile of reasons that puts an end to our questions, something like a purely rational principle or moral law. I don't know of any empirical studies that bear directly on this claim, but it is relevant that there are many Humean philosophers who believe that all reasons are ultimately explained by desires and whose moral convictions do not seem to be undermined by this belief (e.g., Schroeder 2007). Once we distinguish the claim that moral judgments are supported by *some* justifying reasons from the claim that these reasons must be backed up by rational principles, it seems quite plausible to think that people could live with the absence of rational principles that make our moral judgments true. If the argument that there must be such principles if there are to be any reasons at all relies on a claim about our psychology, then the fact that we don't need rational principles would deflate it.

The Kantian, however, would deny that this argument has anything to do with our psychology. What is relevant, she'll say, is what we are like insofar as we are rational; it is only by thinking about what we are like insofar as we are rational beings that we can uncover the metaphysical nature of reasons. As Kant himself puts it, "Since moral laws must hold for every rational being as such, our principles must…be derived from the universal concept of a rational being as such" (Kant 1785/2002, 213/412). The Kantian argument is metaphysical, not empirical: the relevant premise is that a consideration cannot count as a *reason* unless it would be sanctioned by a rational law, and that is just part of what it is for something to be a reason. Indeed, Kant thought it possible that we are not the kind of creature to whom the moral law applies, and on this view it is possible that there are moral reasons that could never be part of the explanation of human action.

True, the Kantian argument is an argument about what a reason is (as opposed to what we take reasons to be as a matter of our psychology). But if we cannot recognize ourselves (even our best, most rational selves) in the description of rational agents in this argument,

then it's not clear why we should care about reasons as Kantians see them. If we are not at all like the reflective creatures Korsgaard and Kant discuss (creatures who seek principled reasons for action that can put an end to our questioning), then the view of reasons they are discussing is not relevant to us. This by itself would not prove that the Kantian conception of a reason is wrong, exactly, but it would make us question what the point of it is.

The Threat to Rationalism: Conclusion

Does the empirical evidence really provide a fundamental challenge to rationalism? What does seem to be threatened is a picture according to which we always arrive at our moral judgments by engaging in rational reflection and we are then motivated to act on these judgments by the sheer recognition of their rational status. It's unlikely that Kant held this extreme view. Whether he did or not, it seems to me that the most important Kantian assumptions about reasoning are compatible with much of the empirical research, because Kantians could be satisfied with a limited causal role for reasoning. Indeed, Kantians could even admit that emotions play an important role in producing moral judgments, because this is compatible with thinking that reasoning is how we *justify* our moral judgments and that rational principles are at the foundation of these justifications. As long as we can reject an emotionally caused but unjustified judgment, the Kantian view would not be endangered. Further, as long as reasoning can succeed in justifying our moral judgments, reasoning needn't *always* have this purpose.

Whether Kantian reasoning can justify our moral judgments depends on some deep issues in metaethics. In particular, it depends on whether there really are any rational principles that provide a foundation for our moral reasons. This is one of the fundamental philosophical disputes between the sentimentalist and the rationalist, and skepticism about the existence of principles of practical reason that have the authority to justify our moral judgments is my own reason for not being a Kantian. But this debate ultimately concerns philosophical questions about the nature of practical reasons, not psychological facts about the causes of moral judgment. As far as the empirical challenge goes, the door for rationalism is still open.

Normative Ethics: Empirical Challenges to Kantian Normative Positions

Again, we will have to take for granted some familiarity with Kantian normative theory, whose core idea is that an action is morally permissible if its maxim (the underlying intention or subjective principle of an action) conforms to the categorical imperative. Two formulations of the categorical imperative have been particularly fruitful: the universal-law formulation, which tells us to act only on that maxim we could will to be a universal law, and the end-in-itself formulation, which enjoins us to treat people never merely as a means but always as ends in themselves.

Reasoning Does Not Lead to Kantian Normative Conclusions

Because Kantianism looks to the intentions underlying our actions to determine their moral status, it stands in contrast to consequentialist theories, according to which the amount of good or value produced by an action is what determines its rightness. This

feature of Kantian normative theory entails that it is impermissible to promote the greater good if the maxim of the good-promoting action is one that cannot be universalized or one that fails to respect rational nature. It is this feature of Kantian normative theory that Joshua Greene, a neuroscientist and philosopher, attacks.

Greene (2014) argues that "characteristically deontological moral judgments" (such as the judgment that it is impermissible to sacrifice one person's life for the sake of saving the lives of several others) are emotional rather than rational. He argues that different moral judgments about cases (intuitions) are caused in different ways, and that, together with some assumptions about when different mental processes are reliable and when not, we have good reason to discount some of our moral intuitions, in particular, the Kantian ones.

To understand Greene's argument, we must confront the trolley (Foot 1967/2002; Thomson 1976). Consider the following cases:

- *Footbridge:* A trolley is hurtling down the tracks. There are five innocent people on the track ahead, and they will be killed if the trolley continues going straight. You are an innocent bystander (i.e., not an employee of the railroad, etc.) standing next to a large man on a footbridge spanning the tracks. The only way to save the five people is to push this man off the footbridge and into the path of the trolley. What would you do?
- *Footbridge Switch:* A trolley is hurtling down the tracks. There are five innocent people on the track ahead, and they will be killed if the trolley continues going straight. You are an innocent bystander (i.e., not an employee of the railroad, etc.) standing next to a switch that controls a trap door, which opens on to the tracks. There is a large man on the trap door. The only way to save the five people is to pull the switch, thus dropping the large man into the path of the trolley. What would you do?

The only difference between *Footbridge* and *Footbridge Switch* is that in the first case you have to push a man to his death, whereas in the second case you pull a switch that has the same result. Either way, the man is killed by the train: one will die and five will live (and if you don't act, five will die and one will live). Despite the similarity in the numbers, people tend to feel differently about these cases. Most people (63%) say that it's morally permissible to pull the switch in *Footbridge Switch*, but only 31% think that it is morally permissible to push the large man in *Footbridge* (Greene et al. 2009). Why this difference?

Greene thinks that the difference between *Footbridge* and *Footbridge Switch* can't be explained rationally. Think about it: in one case, you are right next to the man (close enough to touch him), and in the other case, you are a little farther away, but you can still make him fall into the train by pulling a switch. How could the tiny difference of physical distance make the difference between its being morally o.k. to kill him and its being morally wrong to kill him? Assuming that this tiny difference cannot make a real moral difference, instead of trying to explain our intuitions rationally, Greene offers a causal explanation. He claims that the different intuitions in *Footbridge Switch* and *Footbridge* are explained by the fact that we have two different cognitive systems in our brains. In short, one system, which is emotional and automatic, is engaged when thinking of physically touching the man, and it elicits the emotional judgment that we should not push the man into the train. The other, more reflective, system is engaged in response to reading the relatively cold *Footbridge Switch* case; since our emotions are not engaged, this more reflective system elicits the judgment that we should pull the switch in order to save more people. Let's consider this in a little more detail.

The theory that there are these two systems in the brain is called dual-process theory. System 1 is automatic, emotional, and quick. System 2 is controlled, deliberate, and slow. System 2 is what we normally think of as conscious reasoning or "thinking," but both systems produce judgments. Greene analogizes the two cognitive systems to the automatic and manual modes on a camera. If your camera is on automatic, you can take pictures quickly, but you sacrifice quality. If your camera is on manual, you have more flexibility to cope with, say, different lighting conditions, but you sacrifice speed because of the conscious effort required to set things up (Greene 2014). Greene argues that the two processes in dual-process psychology tend toward different kinds of moral judgment: System 1 produces "characteristically deontological" judgments (concerning rules, rights, and duties), while System 2 produces "characteristically consequentialist" judgments (concerning the greatest benefit to the greatest number). The judgment that we should not sacrifice the one to save the lives of the others is taken to be characteristically deontological because it prioritizes the importance of not violating a man's right to life by killing him over maximizing the number of saved lives; the judgment that we should sacrifice the large man for the sake of the others' lives is taken to be characteristically consequentialist.

Using dual-process theory, Greene hypothesized an explanation for why people tend to make different judgments in *Footbridge Switch* and *Footbridge*. First, our automatic, emotional system of judgment will be triggered by the close and personal nature of the action in *Footbridge* (you have to touch the man to push him on to the tracks), and this system will cause us to judge that we should not push the man. Second, in the *Footbridge Switch* case, without any emotional trigger, our rational, calculative system will determine our judgment and we will consider the outcomes more rationally, thus leading us to say that it would be right to pull the switch. Some of the evidence in support of this hypothesis is neuroscientific: researchers can see from fMRI scans that the parts of the brain that are more active when people judge that it would be wrong to push the large man on to the tracks are the parts of the brain that are associated with emotional activity. And in many cases, patients with emotional deficits due to brain injuries are more likely to make consequentialist judgments. (This research gave rise to the fun title "Consequentialists Are Psychopaths" (Schwitzgebel 2011)). Further, when people are given more time to deliberate or are encouraged to reflect, they are more likely to make consequentialist judgments about a case (Greene 2014).

Suppose we accept the description of our psychology put forward by this research: we agree, for the sake of argument, that consequentialist judgments are associated with conscious/controlled reasoning processes, while deontological judgments are associated with automatic/emotional cognitive processes. What should we conclude? Selim Berker, a critic of Greene's work, has argued that there is no bridge from the "is" of dual-process theory to the "ought" of ethics. He argues that the scientific evidence about the causes of our moral judgments is irrelevant to any claims about which judgments are right or wrong (Berker 2009). According to Berker, in order to derive any conclusion about which judgment is trustworthy, we would have to rely on moral intuitions about what sorts of features of the world our judgments *ought* to be sensitive to. Without such assumptions, we couldn't argue that the cool and calculated judgment is *better* than the judgment that is influenced by the emotions caused by the personal nature of the action.

Berker is right. We must make normative assumptions (about what our judgments ought to be sensitive to) in order to derive a normative conclusion (about which judgments we can trust). But Greene does not deny this. Greene argues that the scientific evidence *together*

with normative assumptions about what counts as good judgment support the conclusion that the consequentialist intuitions are better or more reliable. The argument in favor of trusting consequentialist intuitions depends on the assumption that our judgments *should* (normative term) be insensitive to "mere personal force" or "mere special proximity." Judgments that respond to these considerations alone – absent any other relevant considerations – are biased by irrelevant information. The scientific research supports the claim that our nonconsequentialist judgments are responding just to personal contact and proximity. The assumption that these features – the mere facts of proximity – are morally irrelevant is the sought-after extra, moral, premise.

Judith Jarvis Thomson agrees that there is an uncomfortable incongruity between our willingness to flip the switch and our unwillingness to push the large man. "Well, in *Footbridge Switch* I don't have to use both hands!" doesn't seem like much of a moral argument. But instead of concluding that we ought to count both pulling the switch and pushing the large man as the right things to do (as Greene does), Thomson argues that we ought to judge both actions wrong (Thomson 2008). She introduces another case in which *you* are on a side track that bypasses the five innocents stuck on the main track. In this case, you can pull a switch that would divert the train on to your side track, killing you and thereby saving the five innocent people. Sure, it would be nice if you did this, but, Thomson argues, you are not morally required to sacrifice yourself to save the five. Sacrificing yourself to save five others would be heroic or supererogatory (beyond the call of duty), but not required. Further, if you aren't required to sacrifice yourself, then the large man in *Footbridge Switch* isn't required to sacrifice himself to save five, either. By putting him on to the track, you make him do something that he isn't required to do, and something that you yourself would not do. This seems wrong, and it is plausible that it seems so on the grounds of a principle governing obligations of self-sacrifice (rather than merely because of a response to the anticipated personal contact). This new twist on the case makes us think that maybe we ought to make the same judgment in *Footbridge Switch* and *Footbridge*: in neither case does morality require you to sacrifice a person to save five.

Thompson's conclusion is even more anticonsequentialist than the usual intuitions people have about these cases: we should pay even less attention to the cost–benefit analysis than we originally thought. Thomson's example shows that even if it is true that emotional reactions are driving the anticonsequentialist conclusion about *Footbridge*, this does not mean that there isn't a rational explanation for the anticonsequentialist verdict. Thomson's argument raises the stakes, because many will find it unintuitive to think that it is always impermissible to sacrifice one to save five. Nevertheless, some Kantians might want to accept that this is, in fact, the principled solution (or not: see Hill 1992).

The Motive of Duty

According to Kant, the only actions that are morally worthy are actions done from the motive of duty. The motive of duty is the motive of a good will, which is the kind of thing that is always good to have. It is, Kant thinks, unconditionally good. When we act from duty, we act with the intention – not just an inclination, but the "summoning of every means" – to do the right thing, because it is the right thing, no matter what we feel like doing (Kant 1785/2002, 196/394). A person who acts from the motive of duty does the right thing in virtue of her recognition that morality demands it and her background rational commitment to do what is morally required of her. The motive of duty is important not

only because it is the only morally good motive, but also because morality applies only to beings who are capable of being motivated by duty (i.e., according to Kant, rational beings).

The claim that we could be motivated by a pure sense of duty, unconnected to our desires, has also been the subject of empirical critique. Schroeder, Roskies, and Nichols (2010) argue that given what we know about the causal pathways to voluntary behavior in the brain, it is difficult to sustain the view that we are motivated by our beliefs about what's morally right and wrong. They point out that there is one model for voluntary behavior that is brought about by higher cognitive centers independently of desire, but that this model is of the pathological behavior of Tourette syndrome.

This challenge assumes that to be motivated by duty is to be motivated by a purely cognitive state, such as a belief. Is the motive of duty such a state? Kant sometimes describes the motive of duty as respect for the moral law (Kant 1785/2002, 202/400), and respect is not obviously a belief. As T.M. Scanlon points out, if we understand desires very inclusively, as "pro-attitudes," then there is still room for the motive of duty, because many pro-attitudes, such as "duty, loyalty, or pride, as well as an interest in pleasure or enjoyment," can be brought about by reasoning (1998, 37). If respect is a kind of pro-attitude, and if thinking about the requirements imposed on us by the moral law can cause this attitude of respect (in the way that recognizing beauty in nature can bring about the feeling of awe), then the motive of duty is a pro-attitude that can be brought about by reasoning. This suggestion does not by itself vindicate the motive of duty, but it does provide an avenue of defense for the Kantian on this point. Further defense of the idea would have to show that the recognition of moral requirements can indeed give rise to the right kind of motivational state.

The Threat to Kantian Normative Ethics: Conclusion

Some Kantian ideas about how we ought to behave are threatened by the empirical wave in moral psychology. It is a problem for a principle-based moral theory if there is nothing more to say about why we should behave differently in two similar cases except that we feel differently about the two cases. If we are goal-directed creatures for whom voluntary action is typically acting on our desires, this is a problem for the hypothesis that morally worthy action is motivated by a pure sense of duty that is entirely divorced from our desires. But, as we have seen, there may be more to say about trolley cases that is favorable to a nonconsequentialist picture. And there are other ways of thinking about the motive of duty than as a purely rational force that has no home in goal-directed psychology.

More importantly, however, there are many aspects of Kant's normative theory that are not targeted by the empirical wave. That we should treat each other as fundamentally valuable beings who have our own projects and plans for achieving them; that we should respect other people's judgments about their own lives; that recognizing the worth of other people requires that we spend some of our time helping them pursue their projects and making it possible for them to live autonomously: these are first-order ethical ideas worth taking seriously. One could reject Kant's metaphysical views about rationality, his draconian views about duty, and perhaps even his most dogmatic anticonsequentialist conclusions, but still think he has a lot of important ideas about how we ought to treat other people, morally speaking. It would be a shame if the current wave, which has produced so much valuable work in moral psychology, were to swamp these good ideas.

Acknowledgments

This chapter includes material from my book *Moral Psychology: A Contemporary Introduction* (Tiberius 2014). I thank the publisher for permission to include this material here. I would also like to thank Kelly James Clark, Josh Greene, Tim Schroeder and Connie Rosati for helpful comments and discussion.

References

Aharoni, E., Sinnott-Armstrong, W., and Kiehl, K.A. (2012). Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/conventional Distinction. *Journal of Abnormal Psychology* 121(2): 484.

Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs* 37(4): 293–329.

Bjorklund, F., Haidt, J., and Murphy, S. (2000). Moral Dumbfounding: When Intuition Finds No Reason. *Lund Psychological Reports* 2(1): 29.

Blackburn, S. (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.

Blackburn, S. (1998). *Ruling Passions*. Oxford: Clarendon Press.

Blair, R.J.R. (1995). A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *Cognition* 57(1): 1–29.

Foot, P. (1967/2002). The Problem of Abortion and the Doctrine of Double Effect. In *Virtues and Vices and Other Essays in Moral Philosophy*. Cambridge: Cambridge University Press.

Gibbard, A. (1992). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press.

Greene, J. (2014). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics* 124(4): 695–726.

Greene, J.D., Cushman, F.A., Stewart, L.E., Lowenberg, K., Nystrom, L.E., and Cohen, J.D. (2009). Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment. *Cognition* 111(3): 364–371.

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108(4): 814–834.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. London: Penguin.

Haidt, J. and Bjorklund, F. (2008). Social Intuitionists Answer Six Questions about Morality. *Moral Psychology* 2: 181–217.

Hare, R.D. and Vertommen, H. (2003). *The Hare Psychopathy Checklist-Revised*. North York: Multi-Health Systems.

Hill, T. (1992). Making Exceptions without Abandoning the Principle: Or How a Kantian Might Think about Terrorism. In *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell University Press.

Jones, K. (2006). Metaethics and Emotions Research: A Response to Prinz. *Philosophical Explorations* 9(1): 45–53.

Kant, I. (1785/2002). *Groundwork for the Metaphysics of Morals*, translated by A. Zweig and T.E. Hill. Oxford: Oxford University Press.

Kennett, J. (2006). Do Psychopaths Really Threaten Moral Rationalism? *Philosophical Explorations* 9(1): 69–82.

Kennett, J. and Fine, C. (2008). Will the Real Moral Judgment Please Stand Up? *Ethical Theory and Moral Practice* 12(1): 77–96.

Korsgaard, C.M. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

Lerner, J.S., Goldberg, J.H., and Tetlock, P.E. (1998). Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility. *Personality and Social Psychology Bulletin* 24(6): 563–574.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*, vol. 13. Oxford: Oxford University Press.

Nichols, S. (2010). How Psychopaths Threaten Moral Rationalism: Is It Irrational to be Amoral? In *Moral Psychology: Historical and Contemporary Readings*, edited by Thomas Nadelhoffer, Eddy Nahmias, and Shaun Nichols. Oxford: Blackwell.

Prinz, J. (2006). The Emotional Basis of Moral Judgments. *Philosophical Explorations* 9(1): 29–43.

Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Roskies, A. (2003). Are Ethical Judgments Intrinsically Motivational? Lessons from "Acquired Sociopathy." *Philosophical Psychology* 16(1): 51–66.

Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: The Belknap Press of Harvard University Press.

Schnall, S., Haidt, J., Clore, G.L., and Jordan, A.H. (2008). Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin* 34(8): 1096–1109.

Schroeder, M.A. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.

Schroeder, T., Roskies, A., and Nichols, S. (2010). Moral Motivation. In *The Moral Psychology Handbook*, edited by J. Doris, pp. 72–110. Oxford: Oxford University Press.

Schwitzgebel, E. (2011). Bartels and Pizarro: Consequentialists Are Psychopaths. *The Splintered Mind*. Available from: http://schwitzsplinters.blogspot.com/2011/09/bartels-and-pizarro-consequentialists.html (last accessed July 15, 2015).

Seidel, A. and Prinz, J. (2013). Sound Morality: Irritating and Icky Noises Amplify Judgments in Divergent Moral Domains. *Cognition* 127(1): 1–5.

Sommers, T. (2005). Interview with Jonathan Haidt. *The Believer*. Available from: http://www.believermag.com/issues/200508/?read=interview_haidt (last accessed July 15, 2015).

Thomson, J.J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist* 59(2): 204–217.

Thomson, J.J. (2008). Turning the Trolley. *Philosophy & Public Affairs* 36(4): 359–374.

Tiberius, V. (2013). In Defense of Reflection. *Philosophical Issues* 23(1): 223–243.

Tiberius, V. (2014). *Moral Psychology: A Contemporary Introduction*. London: Routledge.

Wheatley, T. and Haidt, J. (2005). Hypnotic Disgust Makes Moral Judgments More Severe. *Psychological Science* 16(10): 780–784.

25

Naturalism in Metaethics

JUSSI SUIKKANEN

This chapter offers an introduction to naturalist views in contemporary metaethics.[1] Such views attempt to find a place for normative properties (such as goodness and rightness) in the concrete physical world as it is understood by both science and common sense. The chapter begins by introducing simple naturalist conceptual analyses of normative terms. It then explains how these analyses were rejected in the beginning of the 20th century due to G.E. Moore's influential open question argument.

After this, the chapter considers what good general reasons there are for defending naturalism in metaethics. The bulk of the chapter will then survey new semantic and metaphysical forms of naturalism, which in different ways attempt to address Moore's objection. These more recent versions of naturalism – using new resources from philosophy of language, philosophy of mind, metaphysics, philosophy of science, and epistemology – attempt to explain why the open question argument fails.

Naturalist Conceptual Analyses of Normative Terms

The paradigmatic method of philosophical investigation has traditionally been conceptual analysis.[2] When philosophers use this method, they are attempting to capture in simpler terms what a person means when she uses a more complicated term. It is natural to think that what a person means when she says that "Tom is a bachelor" is that Tom is both male and unmarried. Likewise, by calling an animal "a vixen," you mean that it is female and a fox. Conceptual analyses thus attempt to capture the meaning of more complicated terms

1 Metaethics is the study of the nature of normative judgments and normative properties and the meaning of normative language. This is to be contrasted with normative ethics, which studies questions such as which actions are right and wrong.

2 For different ways to understand conceptual analysis and a detailed historical overview of its use throughout the history of philosophy, see Beaney (2014).

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

with definitions that unpack their meaning into simpler terms. These analyses claim, for instance, that a bachelor is by definition an unmarried male and a vixen is by definition a female fox.

Philosophers tend to be interested in terms that are more central parts of our language than "bachelor" and "vixen." One essential part of language which philosophers are interested in includes moral, evaluative, and normative terms. This is the area of language that we use when we discuss what is good and bad; which actions are right and wrong; what we ought to do; what we have reasons to do; what is just, courageous, evil, beautiful, and so on. We also use these terms in our own practical deliberations that lead us to act.

Historically, there have been many attempts to offer conceptual analyses of the central normative terms. It has been suggested, for example, that if you say that "It is good to have a job," what you really mean is that you desire to have a job;[3] this suggested conceptual analysis of the term "good" unpacks the meaning of this analyzed term in terms of the simple idea of what you desire. Likewise, it has been suggested that if I say that an act is "right" what I really mean in simpler terms is that the act brings about more pleasure for everyone than anything else that I could do.[4] According to this analysis, the correct definition of the term "right," which captures its meaning, is "what brings about more pleasure for everyone than anything else an agent could do in her situation."

These simple analyses of "good" and "right" try to explain what these terms mean by talking about only ordinary, easily understandable things like desires and pleasure. Since there is nothing mysterious about human desires or the experience of pleasure, these analyses also explain how we could come to know what is good or right. These analyses furthermore say nothing about God or God's plans. In this sense, it is easy to call these analyses naturalist analyses of "good" and "right," in contrast to supernatural analyses, which would make sense of the meaning of "good" and "right" in terms of, say, God's will.

Moore's Open Question Argument

Despite the attractions of naturalist explanations of the meanings of normative terms, many philosophers are suspicious of them. This line of thought is already present in the work of Richard Price (1723–1791), Ralph Cudworth (1617–1688), Francis Hutcheson (1694–1746), Thomas Reid (1710–1796), and Henry Sidgwick (1838–1900).[5] It culminates in G.E. Moore's open question argument, which Moore formulated in his *Principia Ethica* (Moore 1903/1993, §§ 11–14). In their arguments against naturalist analyses of normative terms, Moore and the other philosophers before him rely on an intuitive standard for when an analysis successfully captures what a speaker means when she uses a given term.

To see how this test works, consider the previous example of the term "vixen." I suggested that when a competent speaker uses this term to describe an animal, what she means is that the animal is female and a fox. Moore observed that if this definition really captures what the speaker means, then the speaker should not be able to begin to consider the question "X is a female fox, but is X a vixen?" This is a closed question for the speaker,

3 Thomas Hobbes is often interpreted to have endorsed this definition in *Leviathan* (see Hobbes 1651/1994, I.6). Whether this was his intention remains controversial.

4 This analysis is often attributed to John Stuart Mill (1871/1998, 59).

5 For references, see Korsgaard (2008, 307, fn. 23). See also Sidgwick (1907, bk. 1, ch. 3).

because she both means "female fox" by the word "vixen" and has already concluded that X is a female fox. Thus, it would make no sense for her to begin to wonder whether the female fox is also a vixen. As a consequence, Moore suggested that successful analyses always lead to closed questions that have the form of the previous question.

Moore then used this test to argue that there cannot be any successful definitions of normative words in naturalist terms. Consider the two analyses previously suggested: "good" means what you desire and "right" means what brings about more pleasure than anything else that you could do. Moore argued that competent speakers can always begin to consider how to answer the following questions: "I desire X, but is X good?" and "X brings about more pleasure than any of the other alternatives, but is X right?" From this observation, he concluded that being what we desire or what brings about more pleasure than other alternatives cannot therefore be what we mean when we use the words "good" and "right." In fact, Moore concluded from these examples that no naturalistic analyses of normative terms could ever capture what the users of these terms mean, because the relevant questions we get from these analyses will always remain open.

Moore's argument was incredibly important for the development of modern metaethics. One reason for this is that Moore concluded from his argument that the normative terms must, as a consequence of the open questions, refer to *non-natural* properties of their own unique kind.[6] With this, Moore single-handedly introduced the standard terminology and dividing lines for naturalism versus non-naturalism debates in metaethics. Non-naturalists defend the Moorean view according to which normative properties are in some fundamental sense of a different kind than the other, more ordinary properties that objects, actions, and states of affairs are usually thought to have. In contrast, naturalists in metaethics think that normative properties are fundamentally of the same kind as other ordinary properties.

Natural and Non-Natural Properties

After Moore's groundbreaking work, it has proven difficult to specify accurately the difference between natural and non-natural properties. There are three main ways to draw the distinction. The first way to draw it is to begin from the observation that we can use our senses to observe which objects have ordinary properties and which don't. We can see the shape of a car, hear how far a bird is from us, and feel how heavy a bag is. The suggestion, then, is that natural properties are all those properties that are empirically observable in this way, whereas whether something has a non-natural property must be known in some other way (Copp 2003).

The second holds that ordinary properties are a part of the causal network of the world. For example, the steepness of a hill causes a ball to roll down and the speed of a bullet causes a lot of damage inside a human body. The claim, then, is that natural properties are all those properties that are a part of these sorts of causal chain. The having of these properties can both be a consequence of other events that take place in the real world and can also cause other things to happen. A defining feature of non-natural properties is that they are in some sense outside the causal network of the physical world (Lewis 1983, § 2).

The third is to argue that natural properties are special because scientists rely on these properties in the best systematic explanations of what happens in the world (Little 1994;

6 For an insightful investigation of Moore's conclusion, see Dreier (2006).

Shafer-Landau 2006, 211). As a consequence, because physics relies on properties such as spin and charge in its most fundamental explanations, these properties are natural properties. The claim, then, is that the properties that are not used in systematic scientific explanations are for this reason of a different kind than natural properties. Such properties are non-natural properties.

These three ways of distinguishing natural properties from non-natural properties have their own problems. They do suggest, however, that at the heart of the disagreement between naturalists and non-naturalists is the question of whether morality is autonomous from empirical and scientific investigation.

I have already explained why G.E. Moore believed that the meaning of normative terms cannot be analyzed in simpler terms which refer to ordinary natural properties. I also mentioned that Moore concluded from the open question argument that unanalyzable normative terms must therefore refer to simple non-natural properties of their own unique kind. Most developments in metaethics during the 20th and the early 21st centuries can be understood as reactions to Moore's open question argument. The advantages of naturalism provided a powerful motive for developing innovative and sophisticated naturalist theories of normative terms and properties that could address Moore's objection to naturalism. Before I introduce these new forms of naturalism, I will first explain the advantages.

General Arguments for Naturalism in Metaethics

There are at least four good reasons for preferring naturalism[7] in metaethics over the Moorean alternatives that are committed to the existence of additional non-natural properties. The strength of these reasons explains in part the ingenuity of the naturalist responses to Moore.

First of all, metaphysical parsimony is an attractive methodological principle in all scientific theorizing.[8] The basic idea of this principle is that we should always look for the simplest theoretical explanations of the events we observe: the kinds of explanation that rely on the fewest possible objects, properties, and laws. Simple explanations have proven to be incredibly powerful, which is why the principle of parsimony has such as good track record. The simple explanations of science help us to predict successfully what happens across a wide variety of different contexts. This has led to the rejection of our reliance on spirits or magic in our explanations and predictions, and it is why we have made genuine intellectual progress.

Because the goal of aiming at metaphysical parsimony has served science so well, we should also rely on this principle in other domains of rational inquiry (including philosophy). Unless it is necessary, we should also avoid positing new kinds of property to explain reality. This is a powerful reason for not adding any non-natural properties to our worldview.

The second reason for preferring naturalism has to do with how we can know about normative properties.[9] As already explained, naturalists believe that normative properties

7 By "naturalism," I mean substantive naturalism: a view about normative concepts and properties. Methodological naturalism, in contrast, is the view that philosophy (ethics included) should proceed by *a posteriori* empirical investigation. See Railton (1989, 155–157).

8 This principle is sometimes called Occam's razor. For an overview of its role and its attractions in philosophical and scientific theorizing, see Baker (2010).

9 For *locus classicus*, see Mackie (1977, 38). For a more recent powerful development of the argument, see (Bedke 2009).

are exactly like other ordinary properties. When it comes to those properties, there is a relatively simple and unproblematic explanation for how we know about them: centrally, ordinary natural properties causally impact our senses. I can know that the book on my desk is rectangular because the photons it reflects hit my retina, which is sensitive to these particles. Through this causal process, I come to see the shape of the book, and through this perception, I reliably come to know that the book is rectangular.

Because naturalists believe that normative properties are like other ordinary properties, they claim that we can know which objects, actions, and states of affairs have these properties by the same empirical methods. For naturalists, normative properties are a part of the causal network of the world, and for this reason naturalists claim that we can observe normative facts as we do all other facts.

If you believe that normative properties are non-natural, however, it will be difficult for you to explain how we come to know about them. Nonnatural properties (1) cannot be observed, (2) are not part of the causal nexus of the world, and (3) are not used in the best scientific explanations. These features of non-natural properties rule out using both our senses and science in knowing when an act is good or right. Non-naturalists must defend *a priori* methods in the epistemology of normative properties. While they often talk about self-evident normative principles that we discover through pure *a priori* reasoning, it is unclear how such reasoning is sensitive to the non-natural properties that non-naturalists believe are part of the furniture of the world.[10]

Third, it is equally difficult for non-naturalists to explain how normative terms came to refer to non-natural properties. In most cases, our words refer to certain objects and substances at least in part because we causally interact with them (Putnam 1975; Kripke 1980). We drink and bathe in a certain liquid, which at some point in human history we named "water." This causal connection explains how the word "water" came to refer to the substance it refers to. In the same way, a certain baby was named "John F. Kennedy." Through the interaction between that baby and the people who gave him his name, that name came to have its reference. The names of substances and individuals, then, refer to specific substances and individuals (rather than to some other substances and individuals) because we are in a causal contact with them when we introduce the names.

If we cannot causally interact with alleged non-natural normative properties, then there will have to be some other explanation for how the word "right" came to refer to the property of rightness (rather than to some other property). With the exception of Ralph Wedgwood's recent work (2007, ch. 4), however, non-naturalists have failed to explain how words come to refer to the relevant non-natural properties. An important advantage of naturalist views in metaethics is that their reliance on causal connections ensures that normative terms refer to the right properties.

The final advantage of naturalism is the failure of non-naturalists to explain how normative properties supervene on natural properties.[11] In order to unpack this abstract idea, we can begin from the thought that it seems obvious that if two actions are identical in all their

10 See Bedke (2009). Devitt (2010) develops general arguments against *a priori* knowledge. For non-naturalist responses to these worries, see, for example, Audi (2005), Shafer-Landau (2003, pt. 5), Wedgwood (2007, pt. 3), and Enoch (2011, ch. 7).

11 This objection was first made by Simon Blackburn (1993a,b). For a non-naturalist response, see Shafer-Landau (2003, 84–89). Frank Jackson has argued that supervenience itself collapses any additional alleged non-natural properties to the base-level natural properties (1998, 122–123). For a critical investigation of this argument, see Suikkanen (2010).

ordinary natural features, they are also the same normatively speaking: equally good and right. However, if normative properties were distinct and additional properties of their own unique kind, it would be difficult to explain why different instances of otherwise identical actions could not have different normative properties. Naturalists, in contrast, have again an easy answer to this challenge. They, after all, claim that normative properties belong to the set of ordinary natural properties, and therefore it is trivially true that there is no difference in the normative properties that two actions have without a difference in their ordinary natural properties.

To summarize, there are a number of powerful reasons for accepting a naturalistic theory of normative terms and properties. The strong intuitions behind Moore's open question argument also give us an idea of what a satisfactory naturalistic view in metaethics would look like. It would need to explain, for example, why it seems that when we talk about what is good or right, we are not merely talking about what we desire or what brings about the most pleasure. How different naturalist views have reacted to this intuition helps us to classify the new forms of naturalism in metaethics. We will first look at naturalist responses to Moore that focus on normative terms. Following this, we will consider views that focus on the nature of the normative properties.

Semantic Forms of Naturalism in Metaethics

In this section, I will introduce different forms of naturalism which attempt both (1) to capture the meaning of normative language in broadly naturalist terms and (2) to address Moore's concern that such analyses fail because they lead to open questions. I will first outline the expressivist response to Moore's open question argument. I will then explain how simple reductive naturalist analyses of normative terms could be defended against Moore's argument by taking into account pragmatic considerations that influence our willingness to use these terms. Finally, I will outline Frank Jackson's network analyses of normative terms.

Expressivism as a Version of Naturalism in Metaethics

Soon after the publication of *Principia Ethica*, philosophers started to consider why we have the intuition that, whatever natural properties an act has, it still remains an open question whether that act is right. One insightful answer to this question came from philosophers who defended early versions of expressivism (Ayer 1936, 106–110; Hare 1952, 94–110).

Expressivists hold that there is a necessary connection between thinking that an act is good and desiring to do that act.[12] Many people intuitively believe that you cannot sincerely conclude that it is good to help other people unless you also have at least some desire to do so. In order to explain this internal connection between normative judgments and motivation, expressivists believe, roughly, that thinking that an act is good consists itself in a desire to do that act (rather than believing that the act has a normative property).[13] Expressivists also claim that the meaning of the sentence "X is good" can be analyzed wholly in terms of the desire-like attitudes for doing X that this sentence conventionally expresses.

12 This thesis is usually called "judgment internalism." For a clear overview of the debates concerning this thesis, see Björklund et al. (2012).

13 For representative recent expressivist texts, see Blackburn (1998), Gibbard (2003), and Schroeder (2008).

Expressivists use this theory to provide an illuminating explanation of why Moorean questions always remain open. When I judge that an act brings about more pleasure than its alternatives, I form a belief of a natural property that the act in question has: I believe that this act has the property of bringing about a lot of pleasure. According to expressivists, judging that an act is good also requires having a desire to do that act. However, no belief about what natural properties an act has can itself settle what I desire. Knowing that it is sunny outside does not move me to go outside, unless I want to be in the sun. This desire is always something distinct from knowing that it is sunny. Forming a desire to do something is therefore something additional that I must always do after I have formed the relevant beliefs about the situation I am in.

According to expressivists, the additional motivation required when using normative terms explains why questions of the form "X is N, but is it good?" always remain open, no matter what natural property N is. When you ask that form of question, you are considering whether you want to do X when you know that X has the property of being N. Whatever natural property N is, thinking about this always makes sense, which is why Moorean questions will always remain open.

Expressivism is a naturalist view, and therefore it shares the general advantages of naturalism (Blackburn 1984; 1998, 49).[14] Expressivists give a fully naturalist explanation of which mental states constitute normative judgments, and this account does not require the existence of any non-natural properties. In fact, all that is required to make sense of normative judgments, according to expressivists, is ordinary, desire-like attitudes, which are commonly recognized in the human sciences, such as psychology and economics.

When metaethicists discuss naturalism, they typically have in mind cognitivist views, according to which normative judgments are ordinary beliefs about natural properties. Naturalism in metaethics, then, is assumed to be a form of cognitivism. But cognitivism is in conflict with noncognitivist expressivism. While expressivists endorse a fully naturalist worldview, they reject the typical cognitivist assumption.

So far, I have only explained what expressivists think about the nature of normative judgments; I have not said anything about what they think about normative properties. The reason for this is both (1) that what normative properties are like is less important from the expressivist perspective and (2) that expressivists disagree about this issue. Some expressivists think that normative properties are metaphysically lightweight "projections" of our desire-like attitudes (Blackburn 1993c, 55–60). While these "shadow properties" are in one sense of their own unique kind (and so "non-natural"), they lack the kind of ontological weight which non-naturalist realists assume that normative properties have.

The standard way to illustrate this view is by analogy to the property of being disgusting. In the order of explanation, the reaction of being disgusted by different objects comes first. When we talk about disgusting things, we project our reaction of disgust on to those things. In the same way, many expressivists think that we react to the world by forming different desire-like attitudes. When we talk about normative properties, we project these reactions on to different actions and states of affairs in the same way as in the case of disgusting things. As a consequence, normative properties are shadows of our desire-like attitudes.

14 In the same way, it should be emphasized that moral error theory too is a form of naturalism. On this view, semantically normative claims presuppose the existence of non-natural properties, but, metaphysically speaking, there are no such properties which make the normative claims false (see Mackie 1977).

Other expressivists think that normative properties are ordinary natural properties (Gibbard 2003, ch. 5).[15] We can again illustrate this view with the example of disgusting things. As a consequence of projecting our reactions of disgust on to wordly objects consistently, there will be a complex natural property that is coextensive with things deemed disgusting. This can be the property of having the property P or the property Q or the property R and so on, where the predicates "P," "Q," "R," and so on describe in detail the natural properties of each and every instance of a disgusting thing. So, roughly, it's the property of being a rotten egg, or a case of incest, or similar. One view of the property of being disgusting, then, is that it is constituted by the underlying coextensive natural property, which can take a form of infinitely long disjunction.

Some expressivists believe that, in the same way, normative terms refer to an underlying, perhaps disjunctive natural property that is coextensive with the set of good things or right things. The property of being right can, on this expressivist view, be roughly the property of being a case of helping an old person across the street, or giving money to charity, or not stealing from someone, and so on. It is just that when we talk about this natural property by using normative terms such as "right," we are expressing our desire toward doing actions that have this natural property. This is something we could not do by merely using ordinary non-normative descriptive terms.

Implicatures and Know-How

In this subsection, I will introduce two attempts to defend simple naturalist reductive analyses of normative terms with the help of recent developments in the philosophy of language. The first view is based on the distinction between what a word means literally (semantics) and what can be done in practice by using that word (pragmatics). The second view is based on the distinction between propositional knowledge and skills.

To illustrate the first distinction, we can use Paul Grice's classic example of a linguistic phenomenon he calls "a conventional implicature" (1989, 46). Consider the following sentences:

(1) Alice is poor but she is honest
(2) Alice is poor and she is honest

The first observation we should make about these two sentences is that they are true in exactly the same situations: both are true when Alice is both poor and honest (it is attractive to understand what a sentence means in terms of the conditions in which it is true).[16] If we accept this simple idea, then we are led to think that the literal meanings of both (1) and (2) are exactly the same.

Despite this, we would not want to utter these sentences in exactly the same situations. The reason for this is that (1) communicates a further thought which (2) doesn't. In practice, the use of "but" in (1) suggests that poor people are not usually honest. The sentence does not say this literally, but if you were to use it in discussion you would communicate

15 Gibbard's argument that there is no logical room over and above the base-level natural properties is similar to Jackson's. See Gibbard (2003, 94–98) and Jackson (1998, 122–123).

16 Truth-conditional semantics culminate in the work of Donald Davidson. See the essays collected in Davidson (1984).

this thought to your audience. This example suggests that many words and utterances imply things that are not part of their literal meaning.

It is possible to use this idea to defend simple naturalist analyses of normative terms against Moore.[17] Let us assume that expressivists are correct about why we think that questions of the form "X is N, but is X good?" are open questions: because uttering the word "good" commits you to desiring X, whereas thinking that X is N does not commit you to this desire, whatever natural property N is. If we accept that words and sentences can imply things that are not part of their literal meaning, then naturalists can argue that using the word "good" commits you to having certain desires because using this word implies that you have them (and not because the relevant desires have anything to do with the literal meaning of the word).

To see how this works, let us return to the claim that "right" means whatever brings about the most pleasure. Why is the question "X brings about most pleasure, but is X right?" an open question? The naturalist response to Moore is that merely describing an act as the one that brings about most pleasure does not commit you to desiring anything, whereas saying that this act is right implies that you want to do X. Accordingly, the utterance that "X is right" would have this implication even if, in saying that X is right, one literally meant that X brings about most pleasure. This then explains why you can hesitate to call an act that brings about most pleasure right even if by saying that it is you mean that it brings about the most pleasure. Saying that the act is right, despite its literal meaning, implies something about your own motivations.

This response to the open question argument leaves the naturalist with one important task. There are many candidates for a naturalist analysis of normative words. "Right" could mean "whatever brings about most pleasure" or "what society accepts" or any number of other things. Naturalists owe us a plausible theory of which one of these analyses we should accept.

One problem is that the practical implicature response to the open question argument means that what we would say about individual cases is not a reliable indication of what we literally mean when we use normative terms. After all, our willingness to use these terms is also influenced by the implicatures these terms carry with them in concrete situations. This is why naturalists must give us an account of how we are to determine what these words mean without relying on our intuitions about individual cases. One suggestion is that we should look at general empirical semantic theories investigated in linguistics. These theories might be able to tell us what literal meanings normative terms must have in order to explain how these terms combine with other linguistic expressions to form meaningful sentences across different contexts.[18]

There is also a second way in which relatively simple naturalist analyses of normative terms have been defended against Moore's open question argument. This second proposal begins from the idea that we do not learn new terms in the form of explicit definitions when we grow up (Smith 1994, 37–38). For example, when I learned to use the word "car"

17 Proposals along these lines have been defended, for example, by David Copp (2001) and Stephen Finlay (2005). Copp believes that normative utterances conventionally implicate desire-like attitudes, whereas according to Finlay the implicatures in question are conversational implicatures.

18 On the basis of this type of consideration, Stephen Finlay has suggested that the literal meaning of "good" in a given sentence can, for example, be analyzed in terms of the promotiveness of an end which is specified by the context of utterance (2014, ch. 2).

correctly, I did not learn this word as a definition of the form "X is a car if and only if X has four wheels, an engine,…" Rather, I become a competent user of this term by acquiring the skill of classifying objects into two categories (those that are cars and those that are not). When I rely on this skill as a speaker, I needn't have any explicit definition in mind, even if I do make use of some implicit standards that constitute my conceptual competence with the word in question.

This idea can be used to construct an attractive response to the open question argument. It can explain why a speaker can begin to consider the question "X is N, but is X good?" even when what she means by saying "X is good" is that X is N (Smith 1994, 36–38). When a speaker utters the sentence "X is good," she is relying on the skill of being able to classify things according to certain implicit standards, which she learned in the form of know-how when she was growing up. What these standards are will not be transparent for the speaker. It can turn out that she is classifying things as "good" by relying implicitly on the standard "Is X N?" In this case, by calling things "good," she would mean that these things are N, even if she might not be aware of this. Likewise, by calling an act "right," a speaker may well be categorizing it as one that brings about the most pleasure even if she is not aware that she is using this standard. This is why Moore's questions of the form "X is N, but is X good?" can remain open even if the definitions from which they are formed correctly capture what the speaker implicitly means.

If naturalists give this response to the open question argument, they still need to explain how we can discover what criteria speakers are implicitly using to classify things, and thus what constitutes the meaning of the terms they use. One suggestion is that, if we let speakers reflect carefully enough on what role the analyzed terms play in their lives and how they use them, they should be able to discover what standards they are relying on implicitly when they use those terms (Smith 1994, 38).

Michael Smith has used this idea to construct a naturalist theory of the meaning of normative terms. According to Smith, one fundamental part of the role of normative terms is that making judgments with these terms creates requirements of rationality (1994, ch. 3). Roughly, on Smith's view, if you think that doing an act is good, then either you desire to do that act or you are being practically irrational.[19] Smith then thinks that when we analyze normative terms like "good," we should find definitions that help us to explain why these terms create the previous type requirements of rationality. Smith concludes that when you say that we have a normative reason to do a certain act in a certain context, what you mean is that more rational (more informed, unified, and coherent) versions of us would want us to do that act in that context (1994, ch. 5). This naturalist analysis of the term "normative reason" might then explain the role which this term has in our deliberations, and might also be something that we can identify as the implicit standard we use in considering what we have reason to do.

Network Analysis

Frank Jackson has argued that, even if we cannot give simple naturalist analyses of normative terms, we can still reductively analyze normative language as a whole in a naturalist fashion (1998, ch. 5). According to this view, we will not be able to find naturalist definitions of the form "To be good is by definition to be N," where "N" is something simple like "what we

19 This thesis is a form of conditional judgment internalism (Björklund et al. 2012, § 2).

desire" or "what brings about most pleasure." Jackson argues that, despite this, we can capture the whole network of normative terms in naturalist terms.

Jackson offers a simple recipe for how the more sophisticated naturalist network analyses of normative terms can be constructed. According to Jackson, normative terms form an interrelated network of concepts. The place a given normative term has in this network determines both (1) the meaning of that term and (2) what a competent speaker must understand about that term (Jackson 1998, 129–130).

At the first stage of Jackson's analysis, we collect all platitudes that, for example, have to do with "right." These trivial claims connect this term to other normative concepts. They might include claims such as:

- If someone judges that X is right, she is disposed to do X.
- If one person says that X is right and another person that it isn't, then at most one of them is correct.
- Right actions in some sense express equal concern and respect to everyone.
- Right actions tend to have good consequences.
- ... (Jackson 1998, 130–131; see also Smith 1994, 39–41)

From this, we can form a very long conjunction of all the platitudes that connect "right" to other normative and non-normative terms. This conjunction will look like this:

(1) (If someone judges that X is right, she is disposed to do X) and (if one person says that X is right and another person that it isn't, then at most one of them is correct) and (right actions in some sense express equal concern and respect to everyone) and (right actions tend to have good consequences) and… (Jackson 1998, 140)

After this, we can replace all instances of normative property words in (1) with variables. As a consequence, we get:

(2) (x, y, z,…) (If someone judges that X is x, she is disposed to do X) and (if one person says that X is x and another person that it isn't, then at most one of them is correct) and (x actions in some sense express y to everyone) and (x actions tend to have z consequences) and…

If we in addition claim that properties x, y, z,… specified by (2) exist and that they exist uniquely, we get the following:[20]

(3) There are (x, y, z,…) and for all ($x^\star$, $y^\star$, $z^\star$,…) [(If someone judges that X is x, she is disposed to do X) and (if one person says that X is x and another person that it isn't, then at most one of them is correct) and (x actions in some sense express y to everyone) and…] and [((if someone judges that X is $x^\star$, she is disposed to do X) and (if one person says that X is $x^\star$ and another person that it isn't, then at most one of them is correct) and ($x^\star$ actions in some sense express $y^\star$ to everyone) and…) if and only if (x is identical with $x^\star$ and y is identical with $y^\star$ and…). (Jackson 1998, 140)

20 This is called a Ramsey sentence, because Frank Ramsey relied on similar sentences in an attempt to represent the empirical content of scientific theories (Ramsey 1978).

This sentence is formulated in purely naturalist terms. It can be used to offer an *a priori* reductivist naturalist analysis of the term "right." According to this analysis, rightness is by definition the unique natural property that is related to other properties exactly in the way that the big conjunction of platitudes tells us that normative properties are related to one another (Smith 1994, 45–46).

The thought behind this complicated analysis is simple. The role that the word "right" plays in our ordinary normative discourse is like a symbol on a map. A map picks out the place of a specific location by specifying how the location is related to all the other locations shown. In the same way, this analysis of the term "right" picks out a certain natural property by specifying how that property is related to other properties in ordinary language and thinking.

This form of naturalism thus enables us to define normative terms as terms that refer to the unique natural properties (whatever they may be) that play the role defined by the place of those terms in the whole network of normative concepts. It is then a matter of further empirical research to discover which particular natural properties the concepts in the whole network of normative concepts refer to. It is true that we could form a relevant open question from the previous definition of "right" which speakers could begin to consider. However, Jackson argues that this definition will only create open questions, because speakers can always be hesitant about whether the discovered analysis based on the network of platitudes is correct. Given how complicated the network analysis is, it is no wonder that it will not be apparent for the speakers that this definition correctly captures what "right" means.[21]

Synthetic Forms of Naturalism in Metaethics

All the naturalist views introduced so far have focused on normative terms and language. They all attempt to show that, in one way or another, normative terms can be defined in non-normative language that refers only to ordinary natural properties. However, there is another alternative. Many metaethicists have recently attempted to naturalize normative properties directly on the level of metaphysics without offering naturalist semantic analyses of normative terms.[22] This new project was made possible by new developments in metaphysics, philosophy of science, and philosophy of language during the second half of the 20th century.

The so-called synthetic naturalists in metaethics think that even if normative terms cannot be defined in simpler naturalist terms, normative properties can still be natural properties. Nonreductive synthetic naturalists believe that normative properties are natural properties of their own, whereas reductive synthetic naturalists think that normative properties are reducible to other natural properties.[23]

21 One objection to Jackson's moral functionalism is that the network of platitudes in the case of normative concepts is not rich enough to pick out unique natural properties as the referents of these terms. For this permutation problem, see Smith (1994, 48–54).

22 These philosophers include Richard Boyd (1988), Nicholas Sturgeon (1988), David Brink (1989), and Mark Schroeder (2007, ch. 4). Because the first philosophers defending the theory worked at Cornell at the time, synthetic forms of naturalism are often called "Cornell realism" in metaethics.

23 Sturgeon (1988, 239–241) is an example of a nonreductivist synthetic naturalist, whereas Schroeder (2007, 61) is a clear example of a reductivist synthetic naturalist.

We can start approaching these views from the idea that, in a sense, the new synthetic forms of naturalism accept the first conclusion of Moore's open question argument (there are no correct definitions of normative terms) but reject the second one (therefore, these terms refer to non-natural properties). When Moore was writing, this combination was not a real option, because it was thought that the meaning of a word determined its reference. If you accept this principle of Frege's philosophy of language (1892/1997, 152), then it is tempting to assume that, if two terms differ in meaning, they must refer to different properties. However, Hilary Putnam (1975) and Saul Kripke (1980) famously argued that there is no reason to accept Frege's principle. They showed that there can be two terms that have different meanings which nonetheless refer to the same individual or substance in all possible worlds. It is easy to use scientific identity statements to illustrate this idea.[24]

According to scientists, water is H_2O molecules. Despite this metaphysical identity, "water" and "H_2O" have different meanings. When you explain the meaning of "water," you will say things like "Water is a transparent, tasteless liquid," "It falls from the sky and fills lakes and rivers," and so on. In contrast, the meaning of the theoretical term "H_2O" has something to do with how hydrogen and oxygen atoms can form larger molecules. Kripke and Putnam's crucial observation was that not everything that is a sample of a transparent, tasteless liquid that falls from the sky and fills lakes and rivers will be water. We can imagine a planet where we find a substance that is like this but is not H_2O (but rather has some other chemical constitution, XYZ). The samples of this substance will intuitively not be water, because we only count H_2O as water. This means that here we have a case where two terms that differ in meaning refer to the very same substance in all worlds.

This analogy creates logical room for the new synthetic forms of naturalism in metaethics. These views hold that even if the predicates "good" and "N" differ in meaning, they can still refer to exactly the same natural property N across all worlds.[25] That is, goodness can be identical with the property of being N even if this identity is not based on the meanings of the relevant words. As a result of this possibility, Moore can be right that the meaning of normative words cannot be captured in simpler terms and yet wrong about normative properties being distinct non-natural properties. This is why the traditional open question argument cannot be used as an objection to the new synthetic forms of naturalism.

All of this, of course, leaves synthetic naturalists with the question: Which natural properties exactly do constitute normative properties metaphysically? According to Cornell realists, in the same way that we use the best overall science of our day to determine which substance is water, we should also rely on the best overall ethical theory to determine which natural properties constitute normative properties. If the simple utilitarian ethical theory (right actions maximize the amount of pleasure) turns out to be true according to the best theory, then the natural property of maximizing pleasure constitutes the property of rightness.

How do we know which ethical theory is the best? At this point, Cornell realists are inspired by the epistemological debates in philosophy of science. Quine argued convincingly that individual elements of scientific theories or even complete theories cannot be empirically tested by observations (1969, 79). What the observations show will always

24 Naturalists in metaethics who use this scientific identity analogy include Boyd (1988, 196) and Brink (1989, 157).
25 This view is most clearly defended by Boyd (1988), Brink (1989), and Schroeder (2007). Peter Railton (1986) has suggested that once we find a naturalist property that plays the role specified by an ordinary normative term, we can revise the meaning of that term so that it will mean the same as the non-normative terms that pick out that property.

depend on many background assumptions outside them. This is because, by making corrections somewhere else in your web of beliefs, you can always make your theory fit what you observe. For this reason, Quine thought that we can only test worldviews against observations holistically (in their entirety).

The new forms of synthetic naturalism in metaethics borrow this form of coherentism from the philosophy of science.[26] They state that we should first attempt to seek a wide reflective equilibrium between our intuitions about cases, general moral principles, and, more broadly, common-sense beliefs and scientific theories about the world. The best overall theory we arrive at as a result of this procedure will then tell us which natural properties constitute normative properties.

Finally, the development of the previous type of synthetic naturalism in metaethics led to an important debate about the nature of normative properties. I have already suggested that one important criterion for whether a given property is a natural property is whether the property can play a role in systematic causal explanations. Some people additionally think that we should believe in the existence of only those properties that are indispensable in causal explanations. Gilbert Harman argued against the new forms of naturalism in metaethics by relying on this principle and the claim that normative properties are not needed in the best causal explanations (1977, ch. 1).

Harman compares two cases. In one, a scientist sees a vapor-trail in a cloud chamber and, on the basis of this observation and her background theoretical beliefs, infers, "There goes a proton." Harman claims that, in this case, the best explanation of the scientist's observation must refer to the proton itself, which then is just one reason why we should believe that the proton exists. Without the proton, there would not have been an observation.

In the second case, a group of hoodlums sets a cat on fire, which leads a passer-by to judge, "That is wrong!" As in the previous case, without her background ethical theory of which actions are wrong, the passer-by would not have made this observation. However, Harman argues that there is no need to refer to the property of wrongness at all in the best causal explanation of the moral judgment. All we need to refer to in that explanation is how the passer-by's upbringing has made her disapprove of burning cats. And, because we do not need to refer to the wrongness of the act to explain the observation, according to Harman, there is no reason to believe that there is such a property in the first place.

Nicholas Sturgeon's (1988) response to this challenge played a big role in the development of new forms of naturalism in metaethics. Sturgeon begins with the idea that the challenge creates an interesting objection to the existence of normative properties only if it shows that those properties would not play a causal role even if they existed (or otherwise the objection would beg the question). This is why, when we think about the cat-burning case, we are allowed to assume that there are normative properties. Sturgeon then argues that, if we are allowed to assume that there are normative properties, these can be used in informative causal explanations of our observations and other events.

One example of a good moral explanation is the following: Hitler ordered the deaths of millions of Jews because he was morally depraved (Sturgeon 1988, 234). In this case, the fact of Hitler's depraved nature would be explanatorily irrelevant only if he would have ordered

26 See, for example, Boyd (1988, 191–192) and Brink (1989, ch. 5). Some metaphysicians consider this type commitment to Quinean confirmation-holism to be a defining feature of naturalism (Devitt 2010, 274). The coherentist methodology also fits nicely with the view that normative terms refer to the natural properties that best explain our use of those terms (Boyd 1988, 195).

the deaths of millions of Jews even in the counterfactual situation in which he was not morally depraved. According to Sturgeon, the naturalist view Harman is objecting to is based on the idea that Hitler's moral depravity is constituted by his non-normative natural properties: by his racism, insensitivity to other people's suffering, and so on (1988, 245–246). This means that, if Hitler had not been morally depraved, he would have been a very different kind of person with respect to his non-normative natural properties. In this situation, he probably would not have ordered the deaths of millions of Jews, and therefore it seems like we can use normative properties in informative causal explanations. This is a good reason to believe that these properties do exist and that they are natural properties.[27]

Sturgeon's naturalist defense of normative properties against Harman's explanatory challenge led to a fruitful debate about moral explanations. Harman (1986) responded that the counterfactual Sturgeon used to test the relevant explanations was too permissive. It would mean, for example, that there was no logical room for moral epiphenomenalism: the view that Hitler's non-normative character traits explain both his actions and why he was morally depraved even if his morally depraved nature doesn't explain anything. After all, defenders of this view could not deny that, if Hitler had been morally different, he might not have done what he did. Because moral epiphenomenalism is a consistent view, Harman argues that the truth of the counterfactual "had X not been P, then Y would not have happened" is not sufficient to make property P explanatory. In response to this objection, naturalists have attempted to find other ways in which higher-order normative properties could be necessary for the best causal explanations even if they are not themselves causally efficacious.[28] These views attempt to vindicate the view that normative properties are real natural properties by showing that these properties are indispensable in the best causal explanations in some more sophisticated way than that Sturgeon initially assumed.

Conclusion

In this chapter, I have outlined how the general advantages of naturalist views in metaethics led to the development of new, sophisticated forms of semantic and synthetic naturalism in response to Moore's open question argument. The debates about these new forms of naturalism are one important reason why so much progress has been made in metaethics during the last 50 years or so. Furthermore, at some point toward the end of the 20th century, as a consequence of these developments, naturalism seems to have become the default view in Anglo-American academic metaethics.

More recently, non-naturalism has made a comeback.[29] There are several reasons for this. First, the new forms of naturalism outlined in this chapter continue to be critically debated, because each faces significant theoretical challenges of its own. Second, some non-naturalists continue to argue that naturalist views in metaethics have implausible first-order

27 Sturgeon also uses this argument to motivate his nonreductivist form of naturalism. It is not necessary that the extension of the normative property that does the explanatory work (and is thus natural) can be picked out with non-normative terms that refer to other natural properties.

28 The best example of an attempt to address Harman's further objection is Jackson and Pettit's (1990) notion of program explanations. For a clear overview of the whole debate about moral explanations and naturalism, see Miller (2013, ch. 8).

29 Currently prominent non-naturalists include Russ Shafer-Landau (2003), Ralph Wedgwood (2007), David Enoch (2011), and Derek Parfit (2011).

normative consequences. For example, the so-called particularists claim that these views cannot account for the complicated way in which whether some consideration is a reason for an act depends on its context (Dancy 2004, ch. 4). Others insist that naturalist views cannot explain why we have reason to care about our future suffering independently of what we care about currently (Parfit 2011, chs. 3 and 24–30).

The third challenge to naturalism is based on the moral twin-earth cases (Hare 1952, 146–149; Horgan and Timmons 1990). This objection is based on the idea that naturalists must describe a causal mechanism that fixes the reference of normative terms to the relevant natural properties. We can then imagine cases in which this mechanism fixes the reference of a normative term to different properties in different communities. If naturalism were true, it would follow from this that these communities would thereafter lack shared concepts with which to have normative disagreements. Some people find this awkward consequence a good enough reason to reject naturalism. Finally, it is also argued that, perhaps with the exception of expressivism, the new forms of naturalism are still unable to explain the close connection between normative thinking and motivation (Wedgwood 2007, ch. 3). The debate between naturalists and non-naturalists is therefore unlikely to go away any time soon.

References

Audi, R. (2005). *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.

Ayer, A.J. (1936). *Language, Truth and Logic*. London: Gollancz.

Baker, A. (2010). Simplicity. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/sum2014/entries/simplicity/ (last accessed July 15, 2015).

Beaney, M. (2014). Analysis. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/sum2014/entries/analysis/ (last accessed July 15, 2015).

Bedke, M. (2009). Intuitive Non-Naturalism meets Cosmic Coincidence. *Pacific Philosophical Quarterly* 90: 188–209.

Björklund, F., Björnsson, G., Eriksson, J., Olinder, R.F., and Strandberg, C. (2012). Recent Work on Motivational Internalism. *Analysis* 72: 124–137.

Blackburn, S. (1984). *Spreading the Word*. Oxford: Oxford University Press.

Blackburn, S. (1993a). Moral Realism. In *Essays in Quasi-Realism*, pp. 111–129. Oxford: Oxford University Press.

Blackburn, S. (1993b). Supervenience Revisited. In *Essays in Quasi-Realism*, pp. 111–129. Oxford: Oxford University Press.

Blackburn, S. (1993c). Morals and Modals. In *Essays in Quasi-Realism*, pp. 52–74. Oxford: Oxford University Press.

Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.

Boyd, R. (1988). How to Be a Moral Realist. In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, pp. 181–228. Ithaca: Cornell University Press.

Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Copp, D. (2001). Realist-Expressivism: A Neglected Option for Moral Realism. *Social Philosophy and Policy* 18: 1–43.

Copp, D. (2003). Why Naturalism? *Ethical Theory and Moral Practice* 6: 179–200.
Dancy, J. (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
Devitt, M. (2010). No Place for the A Priori. In *Putting Metaphysics First: Essays on Metaphysics and Epistemology*, pp. 271–291. Oxford: Oxford University Press.
Dreier, J. (2006). Was Moore Moorean? In *Metaethics After Moore*, edited by Terry Horgan and Mark Timmons, pp. 191–208. Oxford: Oxford University Press.
Enoch, D. (2011). *Taking Morality Seriously: A Defence of Robust Realism*. Oxford: Oxford University Press.
Finlay, S. (2005). Value and Implicature. *Philosophers' Imprint* 5: 1–21.
Finlay, S. (2014). *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press.
Frege, G. (1892/1997). On sinn and bedeutung. In *The Frege Reader*, translated by Max Black, edited by Michael Beaney, pp. 151–171. Oxford: Blackwell.
Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
Hare, R. (1952). *The Language of Morals*. Oxford: Oxford University Press.
Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
Harman, G. (1986). Moral Explanations of Natural Facts – Can Moral Claims be Tested against Reality? *Southern Journal of Philosophy* 24(Suppl.): 57–68.
Hobbes, T. (1651/1994). *Leviathan*, edited by E. Curley. Indianapolis: Hackett Publishing.
Horgan, T. and Timmons, M. (1990). New Wave Moral Realism Meets Moral Twin Earth. *Journal of Philosophical Research* 16: 447–465.
Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
Jackson, F. and Pettit, P. (1990). Program Explanation: A General Perspective. *Analysis* 50: 107–117.
Korsgaard, C. (2008). Realism and Constructivism in Twentieth-Century Moral Philosophy. In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, pp. 302–326. Oxford: Oxford University Press.
Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
Lewis, D. (1983). New Work for the Theory of Universals. *Australasian Journal of Philosophy* 61: 343–377.
Little, M. (1994). Moral Realism II: Non-Naturalism. *Philosophical Books* 35: 225–232.
Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin Books.
Mill, J.S. (1871/1998). *Utilitarianism*, edited by Roger Crisp. Oxford: Oxford University Press.
Miller, A. (2013). *Contemporary Metaethics: An Introduction*, 2nd edn. Cambridge: Polity.
Moore, G.E. (1903/1993). *Principia Ethica*, 2nd edn. Cambridge: Cambridge University Press.
Parfit, D. (2011). *On What Matters,* vols. i, ii. Oxford: Oxford University Press.
Putnam, H. (1975). The Meaning of "Meaning." In *Mind, Language, and Reality*, pp. 215–271. Cambridge: Cambridge University Press.
Quine, W.V.O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays*, pp. 69–90. New York: Random House.
Railton, P. (1986). Moral Realism. *Philosophical Review* 95: 163–207.
Railton, P. (1989). Naturalism and Prescriptivity. *Social Philosophy and Policy* 7: 151–174.
Ramsey, F. (1978). Theories. In *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, edited by D.H. Mellow, pp. 101–125. London: Routledge.
Schroeder, M. (2007). *Slaves of Passions*. Oxford: Oxford University Press.
Schroeder, M. (2008). *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.

Shafer-Landau, R. (2006). Ethics as Philosophy: A Defence of Ethical Non-Naturalism. In *Metaethics After Moore*, edited by Terry Horgan and Mark Timmons, pp. 209–232. Oxford: Oxford University Press.

Sidgwick, H. (1907). *The Methods of Ethics*. New York: Macmillan.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Sturgeon, N. (1988). Moral Explanations. In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, pp. 229–255. Ithaca: Cornell University Press.

Suikkanen, J. (2010). Non-Naturalism: The Jackson Challenge. *Oxford Studies in Metaethics* 5: 87–110.

Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.

26

Evolution and Moral Naturalism

RICHARD JOYCE

Introduction

Methodological naturalism is the principle that requires of anything whose existence we acknowledge that it fit (in some manner to be specified) with our naturalistic scientific worldview. Cats and couches pass this test; ghosts and gods do not. What of morality? There are two ways one might seek a "naturalization" of morality. First, one might investigate how moral judgments, moral practices, and moral institutions fit into the scientific worldview. One might call this the "psychology and anthropology of morality." Second, one might investigate how moral properties – like goodness, evil, praiseworthiness, and so forth – fit into the scientific worldview. That these are very different types of inquiry is best brought out by analogy: it is one thing to try to understand, in scientifically respectable terms, why people believe in ghosts; it is quite another thing to try to understand *ghosts* in scientifically respectable terms.

Methodological naturalists are likely to think that we are well down the road to providing the former kind of naturalization of morality, though many controversies and puzzles remain. But methodological naturalists are considerably less likely to think that the second kind of naturalization of morality is forthcoming. Many of them think that moral properties really are analogous to ghosts: entities that by their very nature do not fit into the scientific worldview. A methodological naturalist, encountering the idea of entities that by nature are non-natural, has straightforward advice: "Don't believe in them." This denial of belief can take different forms. The *error theorist* thinks that the discourse employing these concepts embroils us in falsehood, and that we should henceforth disbelieve and cease to assert propositions that imply that the entities in question exist. (The atheist is a familiar kind of error theorist about theistic discourse.) By comparison, the *noncognitivist* takes a step back and denies that the discourse under scrutiny was ever really committed to the existence of *entities* at all. While the sentence "Stealing is wrong" appears to ascribe the property of wrongness to stealing, noncognitivists hold that appearances are deceptive; they might hold that this judgment amounts to nothing more than "Boo to stealing!" – in which case, asking whether the property of wrongness fits into the scientific worldview betrays conceptual confusion. Both the error theorist and the noncognitivist may be methodological naturalists,

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

and both are likely to think that the former kind of moral naturalization is desirable and forthcoming, but neither of them thinks that a naturalization of moral *properties* (the second kind of naturalization) is possible.

Not all methodological naturalists, however, take this attitude toward morality; many think that the second kind of naturalization of morality is worth pursuing and is forthcoming. The *moral naturalist* holds that moral properties – like goodness, evil, praiseworthiness, and so forth – do fit within the scientific worldview. In order to have a simple version of moral naturalism to hand, consider utilitarianism. A certain kind of utilitarian holds that the only thing of intrinsic moral value is *happiness*, and maintains that we are thus morally obligated to produce the maximal amount of happiness. According to this view, moral properties are identical to or supervene upon the property *being productive of happiness*, which is a causal/psychological quality that, it seems reasonable to assume, fits quite smoothly into the worldview provided by science.[1]

Metaethicists have long argued over these matters. Do moral judgments ascribe properties (as the noncognitivist denies)? If so, do these properties exist (as the error theorist denies)? If so, are these properties of a naturalistic order (as the moral naturalist affirms)? Arguments for and against these metaethical positions are numerous, and the progress remarkable only in its unimpressiveness. By contrast, there are grounds for optimism regarding the first kind of moral naturalization: the psychology and anthropology of morality. After all, questions about the nature of moral judgments, practices, and institutions seem tractable through familiar empirical means, enlisting the resources of such disciplines as psychology, sociology, anthropology, neuroscience, primatology, and economics. Given the progress one might reasonably hope for regarding the naturalization of moral judgments, an intriguing question arises: Could this progress contribute to the other kind of moral naturalization – naturalizing moral properties? Could understanding what moral judgments *are* (why we make them, where they come from, etc.) help establish whether the things that the judgments are *about* – moral goodness, evil, praiseworthiness, and so forth – actually exist? In particular, could understanding the origin of human moral judgment have metaethical implications? Even more particularly, could evidence that human morality is the product of Darwinian selection provide a premise in an argument establishing (or refuting) moral naturalism? It is the goal of the rest of this chapter to explore these questions.

The Evolution of Morality

Humans are moral creatures. In this context, this doesn't mean that humans are morally praiseworthy or admirable (or, for that matter, blameworthy or iniquitous); nor does it mean that humans are proper subjects of moral concern. Rather, it means that humans make moral judgments: We classify our world in terms of moral values, our actions in terms of moral rules, our character traits in terms of moral virtues and vices, and so forth. Where does this way of thinking come from? One may pose this question synchronically, and it is the job of moral psychology to provide an answer – to reveal what faculties are involved in moral judgment. One may also ask the question diachronically, prompting an investigation

1 The moral naturalist need not be a methodological naturalist. She may be willing to countenance the existence of non-natural entities in her ontology; she just happens to think that moral properties are of a naturalistic kind. A utilitarian theist, perhaps?

of the processes by which humans came to make moral judgment in the first place. Recent years have seen a burgeoning of discussion about the evolutionary origins of the human moral faculty.[2] One possibility is that moral judgment is a relatively recent cultural invention, exploiting various psychological faculties that evolved for other purposes. Another possibility is that there exist in the human mind mechanisms that evolved specifically to make moral judgment possible. On the latter hypothesis – which can be called "moral nativism" – a faculty for making moral judgments is a biological adaptation that emerged because this way of thinking provided our ancestors with some sort of reproductive advantage over their competition. What sort of advantage? On this point, hypotheses diverge, but what is striking about them is that they seem entirely compatible with the error-theoretic stance; they do not appear to imply or presuppose that any of our ancestors' moral judgments were *true*. Let me explain.

Most biological traits have nothing to do with truth. It makes no sense to say that one's gall bladder, or any of its activities, is true or false. (Of course, *that one has* a gall bladder is true or false, but that's a different matter entirely.) But *judgments* can be true or false, and thus any evolved psychological faculty designed to produce a kind of judgment does have something to do with truth. It might be argued that the moral faculty (if there is such a thing) governs only feelings and emotions, and that the "judgments" it produces are not really the right kind of item to be assessed as true or false. But this extreme view is fairly implausible upon examination. It can be granted that the moral faculty has a great deal to do with emotions (like anger and guilt), but that is no reason to leap to the conclusion that truth-evaluable judgments have no place in its operations. It is very hard to see, for example, how a mere *feeling*, absent any truth-evaluable judgments, could count literally as the emotion of guilt, for guilt necessarily involves thoughts along the lines of "I have transgressed." Let us assume, then, that the moral faculty produces truth-evaluable judgments (let us, in other words, put noncognitivism to one side). However, this is not to assume that the evolutionary function of the moral faculty is to *track the truth*. The evolutionary function of a trait is the reason that it was selected for; it reflects why it was reproductively useful to our ancestors. In many cases, of course, truth is useful. Our evolved perceptual faculties, for example, are designed to give us an approximately true representation of where objects are in relation to us. There would have been no adaptive advantage for organisms to have a perceptual faculty producing false beliefs about, say, the location of food or the distance of predators. Similarly, if humans have an evolved faculty for simple arithmetic (see Dahaene 1997; Butterworth 1999), the explanation of why such an ability was useful to our ancestors presupposes the truth of the arithmetical beliefs. Again: having false beliefs about how many of your children are accounted for around the campfire, or how many lions are still chasing you after a couple have quit the chase, and so on, is likely to have been disastrous.

But it doesn't always work out this way; sometimes falsehood is useful. For example, people robustly judge themselves better than average in all sorts of ways, including supposing themselves to have an above-average ability to resist the temptation to make unrealistic positive self-evaluations (Friedrich 1996; Pronin et al. 2002). Such everyday delusions might enhance physical health or motivate confident participation in social activities, and thus it is not implausible to suppose that humans have been hard-wired by natural selection to systematically make such unrealistic self-evaluations (see Taylor and Brown 1988; McKay and Dennett 2009). But the beliefs don't need to be true in order to accomplish such adaptive

2 See Krebs (2005), Joyce (2006), Machery and Mallon (2010), Mikhail (2011), and Kitcher (2011).

ends; indeed, a great many of them must be false, since not everyone can be better than average. We must conclude that if there is an innate mechanism producing these kinds of beliefs, it does not have the function of tracking the truth: it exists not because it produces accurate self-appraisals, but rather in virtue of producing self-appraisals that benefit the agent's physical and/or psychological well-being.

Most nativist hypotheses suggest that moral thinking evolved because it played a vital role in enhancing social cohesion. An individual who judges cheating her comrades to be morally repugnant is less likely to do so – perhaps even less likely than someone who sees that cheating will harm her own long-term interests – and thus, on the assumption that cheating is frequently maladaptive, the moral judgment may be selected for. The plausibility of this hypothesis seems independent of whether cheating one's comrades (or anything else) actually *is* morally repugnant. Other moral nativists emphasize the role that moral judgments can play in signaling one's commitment to social projects (Miller 2007; Nesse 2007). Abiding by moral norms frequently involves foregoing immediate profit, meaning that morality can function as a *costly* signaling device. Costly signals correlate with *honest* signals, since the profits that can be gained by giving a dishonest signal will cease to provide a net gain if the signal is sufficiently expensive to produce (Noë 2001). Thus, if one's reproductive capacities depend on being chosen as a partner in various cooperative ventures (hunting, raising a family, etc.), and those doing the choosing will prefer those who are strongly committed to such ventures, then it may be adaptive to advertise one's prosocial allegiance in a costly fashion. Hence, making moral judgments in a sincere manner may be adaptive as a signaling device. Yet, again, there is no pressure to assume that the moral judgments need be *true* in order to play this adaptive role; one might endorse this hypothesis while maintaining an error-theoretic metaethical stance.

Here is not the place to assess the evidence regarding whether moral nativism is true or probable. The jury will be out on this matter for some time yet, and at present no one should be pressing claims either in favor of or against moral nativism with any great confidence. The main point I've stressed is that on all of the live versions of the moral nativist hypothesis, mention of *truth-tracking* is noticeably absent. Let us, then, assume for the sake of argument that one such hypothesis is true; will this have implications for moral naturalism or for metaethics more generally?

The Case for Moral Nativism Undermining Moral Naturalism

Recent years have seen a flurry of discussion surrounding what are called "evolutionary debunking arguments."[3] The basic idea is that discoveries about the evolutionary origins of some familiar mode of thinking might in some manner undermine that thinking. Nobody thinks this applies generally to any evolved psychological faculty; we've already seen that some evolutionary accounts will be *vindicating* (i.e., the cases of perceptual beliefs and arithmetic beliefs). The modes of thinking that seem particularly vulnerable to these arguments are those regarding which an evolutionary genealogy reveals the mechanisms in question not to be truth-tracking. Of course, the argument is nothing so simple (and fallacious) as

3 See Enoch (2010), White (2010), Wielenberg (2010), Brosnan (2011), Kahane (2011), Joyce (2013a; 2016), Fraser (2014), and Kelly (2014).

saying that the faculty in question is not truth-tracking, therefore the judgments it produces are not true. The notion of "truth-tracking" in play here is one that pertains only to the evolutionary function of the faculty, and observing that a psychological faculty does not have the evolutionary function of producing truths does not imply that it *fails* to produce truths. (Human bipedalism does not have the evolutionary function of allowing us to ride bicycles, but it nevertheless allows us to do so.)

Evolutionary debunking arguments vary in the strength of their conclusions, for when we say that such arguments aim to undermine morality, the term "undermine" is intentionally indeterminate. Sharon Street (2006), for example, argues that moral nativism reveals moral realism to be probably false. Moral realism is the metaethical view that moral properties exist as objective features of the world.[4] Street argues that the moral realist, confronted with evidence confirming moral nativism, faces a dilemma concerning the relation between our moral judgments (products of the contingencies of our evolutionary ancestry) and the supposed realm of objective moral facts. On the one hand, if there is no relation, then it would be an astonishing coincidence if many of our moral judgments were even approximately true – a conclusion supposedly disagreeable to the realist. The problem with the other horn of the dilemma, according to Street, is that it is empirically dubious. I have already noted that the usual nativist hypotheses see the ancestral adaptive payoff of having a moral faculty in terms of enhancing certain cooperative tendencies, not in terms of tracking moral truths. Street thinks this "adaptive link hypothesis" is superior to any truth-tracking hypothesis for three reasons: it is more parsimonious, clearer, and more illuminating of the phenomenon it seeks to explain (2006, 129).

Street's antirealist conclusion might be put as follows: "There are no objective moral facts." Yet she doesn't deny the possibility of moral facts *per se* – they will simply be of a constructivist (i.e., non-objective) nature. For a simple example of what such non-objective normative facts might be like, think of facts about the value of money. That a given disk of metal is worth 10 cents may be a fact, but it is a fact constituted by human activity (approximately, by our collective willingness to treat it as worth 10 cents). By contrast, that a given disk of metal is made up mostly of copper is in no sense constituted by human activity; it is an objective matter. The moral constructivist holds that moral values are, very roughly, like monetary values. But just as there's nothing weird or non-naturalistic about monetary values, so there needn't be about constructed moral values. Thus, Street's arguments do not threaten moral naturalism; she allows that moral properties may well exist and fit smoothly into the scientific world order, so long as we construe them as human-constructed features of our world (Street 2012).

A much more radical kind of undermining – and one that is inimical to moral naturalism – would be the error-theoretic conclusion that there are no moral facts at all. What are the prospects of moral nativism providing support for such a view?

Michael Ruse (1986; 2006; 2009) argues that part of what made moral thinking adaptive for our ancestors was its being imbued with a kind of practical objectivity. Judging norms that enjoin cooperation to be human constructs (albeit highly useful ones) renders them susceptible to practical sabotage: such norms seem "escapable," as if there is no good reason for following them if one can get away with secret transgressions. By contrast, judging norms that enjoin cooperation to be somehow *there*, in the nature of things – as if the world

4 Moral realism cuts across moral naturalism, inasmuch as there are non-naturalist versions of realism (e.g., G.E. Moore's intuitionism) and nonrealist versions of naturalism (e.g., Ronald Milo's constructivism).

comes with obligations written into it, as if people come with rights inbuilt, as if some actions are inherently wrong and some right – can strengthen one's resolve to comply. "It is precisely because we think that morality is more than mere subjective desires," writes Ruse, "that we are led to obey it" (1986, 103).

Thinking along these lines can cast doubt on the viability of the constructivism favored by Street (and others). In order for morality to serve its main purposes, the thinking goes, it must be treated as inescapable, as authoritative, as objective. Indeed, one might go so far as to say that if some kind of objectivity is necessary for the uses to which we put moral norms, then that kind of objectivity is an essential part of morality, conceptually speaking. If a system of human-constructed norms couldn't play many of the practical roles that morality plays, then there are grounds for claiming that it could not count, literally, as a *moral* system at all. This seems to be what Ruse intends when he writes, "Ethics is subjective, but its meaning is objective" (2006, 22) and "[W]hat I want to suggest is that…the *meaning* of morality is that it is objective" (2009, 507).

Suppose, then, that Street's evolutionary argument against moral realism were convincing but that, for reasons like those just sketched, her embrace of constructivism were not. In other words, we would have a good argument to the conclusion "There are no objective moral facts," but we would also have grounds for maintaining "Moral facts are essentially objective." From these premises, the only conclusion to draw would be the error-theoretic one: "There are no moral facts at all." We therefore have the structure of an evolutionary argument against moral naturalism before us; now let us critically assess it.

An important component of Street's argument to debunk moral realism is an appeal to parsimony. Regarding explanations of why moral thinking was beneficial to our ancestors, she argues that the adaptive-link hypothesis is superior to the realist's tracking hypothesis (according to which our ancestors profited from developing accurate representations of a realm of objective moral facts) in part because the former is more parsimonious:

> The tracking account obviously posits something extra that the adaptive link account does not, namely independent evaluative truths (since it is precisely these truths that the tracking account invokes to explain why making certain evaluative judgements rather than others conferred advantages in the struggle to survive and reproduce). (Street 2006, 129)

Ruse's argument also appeals to parsimony. He often uses an analogy to make his point, referring to the spike of interest in séances in Europe in the aftermath of the First World War (Ruse 1986, 256–257; 2006, 22–23; 2009, 504–505). Imagine a grief-stricken mother attending such a séance, during which time she comes to believe that her dead son has spoken to her from beyond the grave. We can explain everything that needs explaining about this belief by reference to psychological and sociological factors; there is no need to suppose that the belief might be *true*. Similarly (Ruse thinks), moral nativism explains everything that needs explaining about why humans judge certain actions to have objective moral status; there is no need to suppose that these judgments might be *true*. To do so would be ontologically profligate.

Street's and Ruse's parsimony arguments question whether moral truths are needed to explain our moral judgments. These can be seen as local instances of a broader type of argument against moral realism, questioning whether moral truths are needed to explain *anything at all*. Is there any phenomenon encountered in the world that remains mysterious and inexplicable unless we introduce moral facts into the picture? In fact, however, it seems

plausible to claim that the broader argument collapses into the local form of argument (focused on the explanation of moral judgments), for we cannot acknowledge an instance of a moral fact being necessary to explain some phenomenon without ourselves making a moral judgment. Consider, for example, the existence of Polish concentration camps in 1944. Seeking an explanation for their existence will lead to reference to (inter alia) decisions made by Hitler. Seeking to push the explanation back further – asking why Hitler made these decisions – one might make reference to Hitler's *depravity*. (This example is from Sturgeon 1985.) Here, then, we seem to have a concrete phenomenon (literally concrete: barracks and buildings), the explanation of whose existence makes reference to (inter alia) a moral property. Yet recognizing Hitler's depravity requires one to make a moral judgment – and then, of that moral judgment, the question can always be pressed: Must moral facts be invoked in order to explain it?

This form of argument against moral facts has been articulated by Gilbert Harman (1977; 1986). Harman's version of the argument doesn't focus on the evolutionary explanation for moral judgment, like Street's or Ruse's, but rather on the process of socialization. For the purposes of a debunking argument, though, this is a distraction; the crucial feature is that moral judgments are taken to issue from a process that appears not to be truth-tracking. Harman, however, doesn't push the argument through to an error-theoretic conclusion. His argument contains a crucial qualification: "In the absence of some sort of naturalistic reduction of moral claims, this kind of explanation [i.e., e.g., one that makes reference to Hitler's depravity] does not make it possible to test moral claims empirically in all the ways in which scientific claims can be empirically tested" (1986, 62). It is the conditional nature of this claim that is important to notice. If "some sort of naturalistic reduction of moral claims" is available, then this threat to the existence of moral facts (or the problem of testing moral principles, to put it in Harman's terms) recedes. In other words, the possibility of a reduction seriously challenges the form of debunking argument relying on an appeal to parsimony. Examples will help clarify this.

Consider trying to explain an avalanche. One explanation – call it the "vernacular" – speaks of melting snow and warm weather. Another – call it the "molecular" – speaks only of molecules of hydrogen and oxygen, thermal energy, and so forth. Suppose our evidence confirms the molecular hypothesis. Would this count as a disconfirmation of the vernacular hypothesis? Would we say, "There is no need to posit *snow* as part of the explanation of the avalanche, for to do so would be ontologically profligate"? Clearly not. Our complete molecular explanation of the avalanche may not mention *snow* at all – it may mention only H_2O arranged in a crystalline lattice – but snow just *is* H_2O arranged in a crystalline lattice. There is a reductive relation between the two.

Similarly, we can explain Hitler's decisions without mentioning *depravity* at all. Delving into his past, we might locate character-forming forces that gave him certain beliefs and attitudes, traumatic events that prompted neurotic outlooks, situations that influenced him to respond in unusual ways, and so on – all of which ultimately led him to think that genocide was a reasonable course of action. But it can be responded that some of these psychological traits described in nonmoral terms just *are* depravity. The moral property of depravity is not some ontological extra, to be excluded by considerations of parsimony; it is there, implicitly, in the accepted psychological explanation.

Similarly again, we might be able to explain why the moral faculty evolved without making reference to any moral facts, but rather by presenting an adaptive-link account couched in terms of moral thinking strengthening our ancestors' motivation to cooperate and so on. But it can be responded that moral facts are implicitly present in the adaptive-link

hypothesis, in virtue of standing in identity or supervenience relations to the items explicitly mentioned (such as the tendency to act cooperatively). Again: the moral properties need not be thought of as ontological extras to be excluded by considerations of parsimony.[5]

Harman captures the point succinctly when he writes, "Even if assumptions about moral facts do not directly help explain observations, it may be that moral facts can be reduced to other sorts of facts and that assumptions about these facts do help explain observations" (1977, 13). So the answers to the questions "Are moral truths needed to explain our moral judgments?" and "Are moral truths needed to explain anything at all?" may both be "No" – and one may, moreover, endorse a strong principle of parsimony that obligates us to banish non-explanatory items from our worldview – but for all this, moral facts may survive. The metaethical view that allows moral facts to survive – that maintains that identity or supervenience relations hold between moral properties and the entities appealed to in the endorsed explanatory hypotheses – is moral naturalism.[6]

We started out this section investigating whether moral nativism might undermine moral naturalism. An argument to this effect was outlined, but it turns out that the availability of a viable moral naturalism would render this argument flawed, and therefore this particular debunking strategy appears to be question-begging. On the other hand, moral nativism provides a degree of support for the moral error theory insofar as it offers an explanation of the source of the massive mistake of which the error theorist accuses ordinary thinking. Of course, the moral naturalist will deny that there is any massive mistake requiring explanation; the point is, though, that error theorists who can provide an explanation of the massive mistake they discern in ordinary thinking are certainly in a dialectically stronger position than error theorists who shrug their shoulders when asked for an explanation. In this respect, at least, moral nativism is a friendly result for the error theorist.

Moral naturalism, we have seen, has the potential to debunk a certain form of debunking argument. But might one go further and claim that moral nativism actually provides support for moral naturalism? I'll address some preliminaries in the following section, before taking up this question in the section after that.

Debunking a Debunking Argument

In responding to moral debunking arguments, several critics have suggested that, even accepting the nativist genealogy, moral judgments might, more or less, track the truth. At this point, we should note that there are two different notions of "truth-tracking" in play: an evolutionary notion and an epistemological notion. The former refers to what a psychological faculty is *supposed to do* (evolutionarily); the latter is often taken to refer to a covariation between a belief and the fact that it represents.[7] Suppose an evolved faculty has the function of producing

5 We can now see what is misleading about Ruse's séance analogy. The grieving mother's belief ("Johnny talked to me from beyond the grave") may be explained entirely in naturalistic terms, but it is wholly implausible that one might take what would be necessary to render this belief *true* (i.e., Johnny's ghostly post-mortem existence) and locate identity relations between these truth-makers and the psychological items appealed to in the naturalistic explanation.

6 Harman himself endorses a kind of naturalistic reductionism, though of a relativistic nature. He concludes that "there is empirical evidence that there are (relational) moral facts" (1977, 132).

7 In fact, I think epistemological truth-tracking is quite difficult to spell out, and the covariation analysis runs into difficulties when beliefs concern necessary truths and necessary falsehoods. See Joyce (2016) for discussion.

judgments of the form "X is P." These representational states might covary robustly with Y's being Q, but we would not on that account say that the faculty tracks the truth or has the function of doing so. Whether the faculty tracks the truth depends on whether the judgments covary with those fact(s) that they represent – in this case, X's being P. Whether the faculty *has the function of* tracking that truth depends on whether success at truth-tracking explains the emergence and persistence (and thus the very existence) of the faculty.

Another way of putting this point is using Elliott Sober's (1984) useful distinction between *selection of* a trait and a trait's being *selected for*. The latter indicates that the trait is the target of selection, in that the nature of the trait plays a causal role in the selective process. The former, by contrast, indicates that the trait is a byproduct of the selective process. (Suppose that colored marbles are dropped through a sieve that allows only the small ones to pass, and suppose that all and only the small marbles happen to be red. There has been selection *of* redness but not selection *for* redness; the sieve selects for smallness.) If a faculty for moral judgment is the product of natural selection, it is possible that, even if its truth-tracking quality was not selected for, there has nevertheless been selection of a truth-tracking trait.

Given these considerations, there are several ways in which critics of debunking arguments may proceed. They may accept that the evolved moral faculty (assuming there is one) is not evolutionarily truth-tracking, but claim that it may be, nonetheless, epistemically truth-tracking. Perhaps the moral facts are causally connected to the natural facts invoked by the moral nativist. Perhaps there is a naturalistic reduction between the two. Perhaps there is a "third factor" that explains both the natural facts in question and the moral facts. Alternatively, the critic may maintain that the moral faculty is (at least in part) truth-tracking in the evolutionary sense. The differences among these options are subtle and, in some cases, potentially problematic; it is not my goal here to tease them apart. But it is reasonable to observe that, despite differences, these critical responses form a family of strategies.

Kevin Brosnan (2011), for example, suggests the possibility that cooperation with others is morally good. The evolutionary process would explain both why we believe that cooperation with others is morally good (because doing so enhances the tendency to cooperate in an adaptive manner, say) and why cooperation is in fact good (because it tends to promote well-being, say). David Enoch (2010) presents a structurally similar argument. He speculates that survival or reproductive success is morally good, and that Darwinian forces have shaped our moral beliefs such that they often concern actions and events that promote survival and reproductive success. Thus, even if the *truth* of our ancestors' beliefs does not figure in the account of why they were adaptive, nevertheless they *were* (sometimes and nonaccidentally) true. Erik Wielenberg advocates another such argument, supposing that natural selection has provided humans with beliefs concerning individuals being surrounded by "a kind of moral barrier that it is…illegitimate for others to cross" (2010, 444–445). Such a belief might well have been adaptive in various ways. Moreover, the very cognitive capacities that make forming such a belief possible also guarantee (or at least probabilify) that one has such a "moral barrier," thus ensuring the belief's truth.

All the aforementioned philosophers aim to find a place for moral facts within the naturalistic nativist hypothesis, though most of them are more concerned with outlining the formal relation between moral and natural facts than they are with making the relation plausible. One diagnosis of why someone might think that a mere sketch might suffice to defeat a debunking argument is a failure to distinguish between debunking arguments that

aim to establish an error theory and debunking arguments that aim to establish the more modest conclusion that our moral judgments lack justification. (Thus far, this chapter has been concerned exclusively with the former. The reason for this is that the thesis that moral judgments lack justification is consistent with either the truth or the falsity of moral naturalism, and thus is peripheral to the focus of this chapter.) If one aims to establish an error theory – that no moral judgments are true – then an opponent's establishing the mere possibility that moral facts implicitly reside in the nativist genealogical account is pertinent. But if one is arguing that moral nativism undermines the epistemic status of our moral judgments, then an opponent's establishing the mere possibility of moral facts implicitly residing in the nativist genealogy misses the point. Analogy: If I deny that life on Mars ever existed, then your convincing me of the possibility that it once existed in underground aquifers should give me pause. By contrast, if my claim is that we have no grounds for believing one way or the other whether life on Mars existed, then your convincing me only of the *possibility* that it once existed in underground aquifers doesn't budge my claim. I may be quite aware of this possibility; my contention is that the evidence is insufficient to grant the belief in Martian life warrant.

It is worth noting at this point that though the previous section explored the possibility of establishing an error theory on the basis of moral nativism, this is by no means the typical debunking strategy. Street is certainly no error theorist, and nor is Harman. I myself have argued for an error theory (Joyce 2001), but on *a priori* metaethical grounds, not via an evolutionary debunking argument. The debunking argument that I have offered (Joyce 2006) has as its conclusion that no moral judgments are justified.[8] Ruse's metaethical position is open to interpretation; though earlier I quoted passages that seem to provide a bridging premise to an error-theoretic conclusion, Ruse never (to my knowledge) explicitly endorses that view. Generally, he expresses himself as (like Street) opposed to moral realism, not opposed to moral facts *per se*. If, then, a critic provides only a sketch of moral naturalism because he or she thinks that this is all that is necessary to refute the error-theoretic debunking argument, it is not at all clear who among actual philosophers is the target.

Another possible diagnosis of why a critic might think that a mere sketch of moral naturalism suffices to refute the debunking argument – even one with an epistemic conclusion – is that she interprets the would-be debunker as claiming that whatever the nature of moral facts, our beliefs concerning them would lack justification. Wielenberg, for example, attributes to the epistemic debunker the position that even if moral facts exist, nativism shows that we would not know of them. He thus grants himself license to offer only a sketch of the nature of moral facts (mentioned previously) – virtually stipulating their existence and nature – and then proceeds to show how knowledge of these facts would be consistent with moral nativism. But this misrepresents the epistemic debunking argument. To recycle an analogy: Suppose one's claim that we do not know one way or the other whether life on Mars ever existed meets with the following response:

> Imagine that Martians live in large underground cities because their leaders want to stay hidden from human view, but suppose also that a small minority of Martians desire to secretly reveal the truth to humans. These rebels leave clues for our little robotic vehicles to find: rocks

8 I once made the ill-thought-through decision to label the thesis that moral judgments lack justification a version of "error theory" (Joyce 2006, 223), but have since recanted this (Joyce 2013a; 2016).

> moved overnight, boulders that look a bit artificial, odd-shaped shadows, and so forth – that is, things that we actually do occasionally observe on the Martian surface. If the facts are as just described, then it can be said that those actual humans who believe in Martian life on the basis of weird-shaped rocks and so forth form their beliefs via a reliable process: these clues connect their beliefs about Martian life with actual Martian life. Therefore, it is possible that we (or at least a discerning few of us) know that there is life on Mars.

This response seems entirely wrongheaded. If one claims that we lack moral knowledge, one's opponent is not free to stipulate moral facts (and the nature of our connection with them) as he or she wishes and then declare that the skeptical claim is mistaken.

In short, merely pointing out how some form of moral naturalism *might* block a debunking argument does not show that any form of moral naturalism *does* block a debunking argument. In order to actually debunk a debunking argument, a particular theory of moral naturalism has to be carefully articulated and defended from criticism. Possibilities have to be converted into plausibilities.

The Case for Moral Nativism Supporting Moral Naturalism

One of the more developed versions of evolutionary moral naturalism currently available is that put forward by Kim Sterelny and Ben Fraser (forthcoming). Sterelny and Fraser argue that the evolutionary function of moral cognition is, in part, to track moral facts. More particularly, they argue that moral cognition (when functioning properly) is sensitive to "facts about profitable forms of cooperation, about social arrangements and cognitive dispositions positively and negatively relevant to the stable exploitation of those opportunities" (forthcoming). "Given a social and physical environment, and a set of interacting agents with their opinions and motives," they write, "there will be facts about whether their current norms are efficient means to stable and profitable cooperation" (forthcoming). These facts are the naturalistic reductive base of moral facts; they are what human moral cognitions are designed to pick out. This allows that one culture's moral system may be closer to the truth than another's; in principle, a culture's moral views (assuming homogeneity) could be entirely true. These moral judgments, moreover, will be true in virtue of the obtaining of certain naturalistic facts – complex and epistemically hard-to-access facts, to be sure, but certainly facts that fall under the purview of the natural sciences.

Sterelny and Fraser invoke the "Canberra Plan" to support their evolutionary naturalism (see Lewis 1970; Jackson and Pettit 1995; Jackson 1997). This method (popular among some prominent philosophers connected to the Australian National University in Canberra) consists of reconstructing the folk platitudes surrounding a concept – a reconstruction that may involve systematization and reflective equilibrium – the results of which can then be assessed from the perspective of our best scientific worldview. They suggest, furthermore, that the folk platitudes in question need not all be items of propositional knowledge, but may include practical skills. Thus, that the folk can successfully *use* a set of concepts for negotiating the world stands in favor of those concepts being vindicated by the Canberra Plan, as opposed to being debunked. Ancient astronomy, for example, for all its horrendous errors, was to some extent responsive to celestial facts in a manner that "guided navigation, calendar construction and time keeping" (Sterelny and Fraser forthcoming), and thus this methodology partially

vindicates its conceptual scheme; we are not error theorists about the ancient discourse of *stars*, *the moon*, and so on. The concept *witch*, by contrast, does not pass:

> Even if those persecuted [as witches] were an identifiable subgroup…discrimination did not leverage adaptive behavior, even by the lights of the witch-burners. It did not prevent crop failures or other misfortunes…So nothing in the world remotely corresponds to the witch-identifying maxims; nor did witch representation leverage adaptive behaviour. (Sterelny and Fraser forthcoming)

An application of the Canberra Plan, Sterelny and Fraser think, will underwrite the reductive relation they propose between moral concepts and the cooperation-oriented naturalistic properties in question, thus offering a partial vindication. They recognize that the world may not supply a property that perfectly satisfies the best systematization of folk moral platitudes, but they think that, on balance, the kind of naturalistic properties they have in mind are the best deservers of this honor. The folk moral concepts may contain errors, but if they also contain truths and, importantly, if they sufficiently promote adaptive behaviors, then they may pass the Canberra test. This is where the nativist hypothesis may provide specific support for moral naturalism, since nativism implies that moral cognition was, relative to ancestral environments, broadly adaptive.

That Sterelny and Fraser present a nuanced and noteworthy form of evolutionary moral naturalism is not in doubt; what is less clear is whether they offer anything to counter the standard kinds of objection that are typically leveled at moral naturalism. If their aim is to make a move against a presupposed backdrop of sympathy with moral naturalism, then this is no great criticism; but if their theory is supposed to be persuasive in a broader metaethical setting – one which includes interlocutors with grave misgivings about moral naturalism in general – then it is doubtful that they make very much headway. Of the several "grave misgivings" one might have about moral naturalism, its inability to adequately accommodate any notion of real normativity is the one on which I will focus in my closing discussion. (For further criticism of moral naturalism, see Mackie 1977; Horgan and Timmons 1991; Kelly 2004; Parfit 2011, pt. 6; Tropman 2012; Miller 2013.)

The moral wrongness of an action, it might be said, directs one against its performance; it tells one what to do. But natural properties appear to be inert in this respect. The fact that some action promotes stable forms of cooperation does not, in and of itself, guide one to perform it. The key phrase here is "in and of itself." Of course, if a person cares about promoting cooperation, then that an action has this property will be of practical import. But for the person who lacks this care, the fact that an action has this property may be an item of little interest; indeed, it looks like something that may be (*ceteris paribus*) legitimately ignored by him or her.

The view often promoted by reductive naturalists (explicitly or otherwise) is that a given property might be picked out by two different terms: a purely descriptive predicate (e.g., "…erodes stable cooperative practices") and a moral predicate (e.g., "…is wrong") – the latter of which may also have certain extra features, such as expressing to one's audience, by linguistic convention, one's disapproval. But in embracing such a view, the naturalist locates normativity in those who make moral judgments: in the force with which they utter their judgments, in the emotions they thereby express, in the persuasive intentions that accompany their utterances, and so forth. It may be objected, however, that a moral realism deserving of the name should locate normativity *in the object*. It is surely

supposed to be the wrongness of stealing that should guide one not to do it, not the qualities of the utterance made by the person telling one not to steal (see Dancy 2006). Moral naturalists cannot accommodate this. The best they can manage is to locate a naturalistic property that most people, as a matter of contingent fact, care about, implying that most people, as a matter of contingent fact, will have reason to comply with moral imperatives and promote moral values.[9]

Many find this aspect of moral naturalism wholly inadequate, and it is not unreasonable to think that it represents the failure to accommodate a central folk moral platitude. The folk think that what Jack the Ripper did was *wrong* regardless of whether he cared about morality. Indeed, his not giving a fig wouldn't incline us to concede that morality is of no practical significance to him – that he can legitimately ignore it – but rather would encourage us to condemn him all the more vigorously. (And the fact the Jack went to his grave unpunished eliminates any complications that might arise concerning the punitive responses of his fellows, yet in no way softens our moral condemnation.) This failure to accommodate so central a platitude suggests that no set of naturalistic properties could *be* wrongness.

Sterelny and Fraser may counter that it is one of the great virtues of the Canberra Plan that concepts can survive the revelation that the folk are mistaken about certain elements of the concept. Perhaps the folk are simply misguided about the idea that wrongness in and of itself should guide practical choices; perhaps they are misguided in thinking that moral facts are relevant to agents irrespective of their desires or interests. But the flexibility of the Canberra Plan cuts both ways. Perhaps instead the folk are simply misguided in thinking that moral properties exist at all. After all, there is nothing inherent in this method that stipulates that only naturalistic and scientifically respectable platitudes count, or that they count for more; if the folk tacitly embrace platitudes that are inimical to moral naturalism, then so be it. One should expect to find many folk concepts for which a Canberra Plan-type naturalization will fail, for there is no reason to think (and many reasons to deny) that folk concepts emerged and evolved with methodological naturalism as a background constraint. J.L. Mackie, for example, argues that the fairly unattractive kind of non-naturalism espoused by G.E. Moore (1903/1948) does a pretty good job of capturing the folk conception (Mackie 1977, 31–32). Nor, it is important to note, is there anything inherently conservative about the Canberra Plan; it is no more a tool for showing that things exist than it is a tool for showing that things do not exist; it should be considered no more friendly to naturalists and realists than it is to error theorists.

What of the supposition that moral cognition "leverages adaptive behavior" – does this lean the argument back in favor of the naturalist? This consideration, upon reflection, is less decisive than appearances suggest. For a start, it is not obvious that it divides examples cleanly in the intuitive manner. Sterelny and Fraser maintain that the concept *witch* failed pragmatically: burning "witches" didn't prevent crop failure or other misfortunes. But it is conceivable that the concept may have been useful in other ways: encouraging social homogeneity and reducing instances of defection, for instance. Certainly it is conceivable that the

9 Note that the issue here does not, in the first instance, concern *motivation*. The claim is not that buried in folk thinking is motivation internalism: the thesis that moral judgments necessarily motivate. Sterelny and Fraser admit that according to their view, moral facts are such that "their power to motivate us is contingent" (forthcoming). This admission, however, does not speak to the complaint under discussion: that moral naturalism fails to accommodate any real normativity.

use of the concept could be adaptive (even if at the actual world it generally was not), which seems to imply that, by Sterelny and Fraser's lights, the concept fails only contingently. The concern is not, of course, that there are possible worlds at which witches exist (presumably there are); the concern is that there are fairly nearby possible worlds at which the use of the concept *witch* is all-things-considered quite useful – leading to the unsavory conclusion that witches thereby exist at those possible worlds.

Or consider another example: phlogiston. This obsolete concept, posited by 17th-century scientists to explain combustion and rusting, could be employed perfectly well in everyday contexts. A person could divide materials according to whether they were phlogisticated or not (wood = yes; water = no), could select only the former type to build the household fire, could seek out escaping phlogiston on cold winter evenings, could point to flames and say, "See the phlogiston escaping," and never be caught out, and so forth. Only in careful experimental conditions do the theory's flaws become apparent (e.g., measuring the mass of items before and after burning). This everyday usefulness of the concept *phlogiston*, however, is clearly not enough to outweigh its central erroneous platitude about combustion being a process of a stored substance released from flammable material. There is no phlogiston.

In the case of moral cognition, the "adaptive leverage" can be overstated. Certainly moral nativism implies that moral concepts must have pulled their weight in terms of enhancing reproductive fitness. But that was back in the Pleistocene. In the modern era of world wars and terrorism, the usefulness of morality can be questioned (see Moeller 2009; Garner 2010; Marks 2013). Of the hundreds of thousands who died in the trenches of the First World War, for example, how many would have been there had they not been susceptible to the storm of moralistic propaganda surrounding recruitment, or the sense of dutiful camaraderie felt toward their fellow soldiers? Would large-scale wars or terrorist violence even occur without the possibility of morality's being recruited to their service?

These are rhetorical questions, but let us suppose that we were to come to the conclusion that while moral cognition *was* adaptive for our ancestors, it *is*, all things considered, no longer useful. This draws attention to the possibility that the folk moral platitudes, considered only as they currently stand, might fail the Canberra test, whereas the folk moral platitudes of our prehistoric ancestors might pass. Such an odd conclusion could be avoided by emphasizing that the current concepts are the natural continuers of the past concepts, and therefore presumably inherit their referential success. Nevertheless, an uncomfortable *contingency* lingers. A concept that is put to good use in one environment may be futile in another not-so-different environment, and if this adaptive utility is what tips the scales between a Canberra-style vindication versus a debunking, then it looks like the wrong kind of criteria are being used to determine whether or not moral properties exist.

Conclusion

Whether one looks to moral nativism as a way of debunking morality or as a way of vindicating morality, the arguments discussed in this chapter tend to lead to the same place: to questions about the viability of moral naturalism. If one hoped, therefore, to use such an argument to verify or to refute moral naturalism, then one is likely to be disappointed. It is

not unreasonable to claim, however, that it is the moral naturalist who bears the burden of proof to present a compelling theory. Yet even this somewhat weak declaration may be sufficient to power a debunking argument to an epistemological conclusion (rather than to an error-theoretic conclusion). If (1) moral nativism raises doubt about the truth of our moral judgments and (2) moral naturalism would dispel this doubt, but (3) there are serious doubts surrounding the viability of moral naturalism (whose resolution we await), then it seems sensible to conclude that our confidence in our moral judgments should be provisionally lowered. This is essentially the "more modest" conclusion encountered earlier: that our moral judgments lack justification.

This chapter has discussed the rather exciting idea that empirical findings can impact on metaethics: that philosophers can roll up their sleeves and get their hands dirty using *a posteriori* data to solve perennial problems. But ultimately the arguments examined lead back to the need to do metaethics the old-fashioned way. Sometimes there is nothing for it but to wash your hands, roll your sleeves back down, and sink back into the armchair.

Acknowledgments

This chapter has drawn freely from Joyce (2013a,b; 2016).

References

Brosnan, K. (2011). Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge? *Biology and Philosophy* 26: 51–64.

Butterworth, B. (1999). *What Counts? How Every Brain is Hardwired for Math.* New York: The Free Press.

Dahaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.

Dancy, J. (2006). Nonnaturalism. In *Oxford Handbook of Ethical Theory*, edited by D. Copp, pp. 122–145. Oxford: Oxford University Press.

Enoch, D. (2010). The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It. *Philosophical Studies* 148: 413–438.

Fraser, B. (2014). Evolutionary Debunking Arguments and the Reliability of Moral Cognition. *Philosophical Studies* 168: 457–473.

Friedrich, J. (1996). On Seeing Oneself as Less Self-Serving than Others: The Ultimate Self-Serving Bias? *Teaching of Psychology* 23: 107–109.

Garner, R. (2010). Abolishing Morality. In *A World Without Values*, edited by R. Joyce and S. Kirchin, pp. 217–233. Dordrecht: Springer Press.

Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.

Harman, G. (1986). Moral Explanations of Natural Facts: Can Moral Claims be Tested against Moral Reality? *Southern Journal of Philosophy* 24(Suppl.): 57–68.

Horgan, T. and Timmons, M. (1991). New Wave Moral Realism Meets Moral Twin Earth. *Journal of Philosophical Research* 16: 447–465.

Jackson, F. (1997). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, F. and Pettit, P. (1995). Moral Functionalism and Moral Motivation. *Philosophical Quarterly* 45: 20–40.

Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.
Joyce, R. (2013a). Irrealism and the Genealogy of Morals. *Ratio* 26: 351–372.
Joyce, R. (2013b). Ethics and Evolution. In *The Blackwell Guide to Ethical Theory*, 2nd edn., edited by H. LaFollette and I. Persson, pp. 123–147. Oxford: Blackwell.
Joyce, R. (2016). Evolution, Truth-Tracking, and Moral Skepticism. In *Essays in Moral Skepticism*. Oxford: Oxford University Press.
Kahane, G. (2011). Evolutionary Debunking Arguments. *Noûs* 45: 103–125.
Kelly, D. (2014). Selective Debunking Arguments, Folk Psychology, and Empirical Moral Psychology. In *Advances in Experimental Moral Psychology: Affect, Character, and Commitments*, edited by J. Wright and H. Sarkissian, pp. 130–147. New York: Continuum.
Kelly, E. (2004). Against Naturalism in Ethics. In *Naturalism in Question*, edited by M. De Caro and D. MacArthur, pp. 259–274. Cambridge, MA: Harvard University Press.
Kitcher, P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.
Krebs, D. (2005). The Evolution of Morality. In *The Handbook of Evolutionary Psychology*, edited by D. Buss, pp. 747–771. New Jersey: John Wiley and Sons.
Lewis, D. (1970). How to Define Theoretical Terms. *Journal of Philosophy* 67: 427–446.
Machery, E. and Mallon, R. (2010). Evolution of Morality. In *The Moral Psychology Handbook*, edited by John M. Doris and the Moral Psychology Research Group, pp. 3–46. Oxford: Oxford University Press.
Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin Books.
Marks, J. (2013). *Ethics Without Morals: In Defence of Amorality*. New York: Routledge.
McKay, R. and Dennett, D. (2009). The Evolution of Misbelief. *Behavioral and Brain Sciences* 32: 493–510.
Mikhail, J. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.
Miller, A. (2013). *Contemporary Metaethics: An Introduction*. Malden: Polity.
Miller, G. (2007). Sexual Selection for Moral Virtues. *Quarterly Review of Biology* 82: 97–121.
Moeller, H.-G. (2009). *The Moral Fool: A Case for Amorality*. New York: Columbia University Press.
Moore, G.E. (1903/1948). *Principia Ethica*. Cambridge: Cambridge University Press.
Nesse, R. (2007). Runaway Social Selection for Displays of Partner Value and Altruism. *Biological Theory* 2: 143–155.
Noë, R. (2001). Biological Markets: Partner Choice as the Driving Force behind the Evolution of Mutualisms. In *Economics in Nature: Social Dilemmas, Mate Choice, and Biological Markets*, edited by R. Noë, J. van Hoof, and P. Hammerstein, pp. 93–118. Cambridge: Cambridge University Press.
Parfit, D. (2011). *On What Matters*. Oxford: Oxford University Press.
Pronin, E., Lin, D., and Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self versus Others. *Personality and Social Psychology Bulletin* 28: 369–381.
Ruse, M. (1986). *Taking Darwin Seriously*. Oxford: Basil Blackwell.
Ruse, M. (2006). Is Darwinian Metaethics Possible (And If It Is, Is It Well-Taken)? In *Evolutionary Ethics and Contemporary Biology*, edited by G. Boniolo and G. de Anna, pp. 13–26. Cambridge: Cambridge University Press.
Ruse, M. (2009). Evolution and Ethics: The Sociobiological Approach. In *Philosophy After Darwin*, edited by M. Ruse, pp. 489–511. Princeton: Princeton University Press.
Sober, E. (1984). *The Nature of Selection*. Cambridge, MA: MIT Press.
Sterelny, K. and Fraser, B. (Forthcoming). Evolution and Moral Realism. *British Journal for the Philosophy of Science*.
Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127: 109–166.

Street, S. (2012). Coming to Terms with Contingency: Humean Constructivism about Practical Reason. In *Constructivism in Practical Philosophy*, edited by J. Lenman and Y. Shemmer, pp. 40–59. Oxford: Oxford University Press.

Sturgeon, N. (1985). Moral Explanations. In *Morality, Reason and Truth*, edited by D. Copp and D. Zimmerman, pp. 49–78. Totowa: Rowman and Allanheld.

Taylor, S. and Brown, J. (1988). Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin* 103: 193–210.

Tropman, E. (2012). Can Cornell Moral Realism Adequately Account for Moral Knowledge? *Theoria* 78: 26–46.

White, R. (2010). You Just Believe that Because… *Philosophical Perspectives* 24: 573–615.

Wielenberg, E. (2010). On the Evolutionary Debunking of Morality. *Ethics* 120: 441–464.

27

Scientific Naturalism and the Explanation of Moral Beliefs: Challenging Evolutionary Debunking

WILLIAM J. FITZPATRICK

Approaches to the Study of Morality: Science and Philosophy

Consider a typical moral judgment:

M: Human trafficking (where vulnerable people are held and pressed into various forms of service against their will) is seriously morally *wrong*, and those engaging in such mistreatment of others *ought* to be stopped and punished.

Such a proposition can be viewed from two very different perspectives. First, you might actually make this judgment yourself, perhaps in response to a news report about trafficking. From this *internal* perspective, you are engaging in the judgment as a committed moral agent, and you will typically have reasons to give in defense of it if challenged. You might even be led to act on this judgment, perhaps by making a contribution to an anti-trafficking organization. Alternatively, you might simply take on the role of an observer, noting that *others* make this judgment and sometimes alter their behavior accordingly, and you might then seek to explain these things. From this *external* perspective, you observe certain empirical phenomena in the world (people's making certain judgments and behaving in certain ways) and go on to inquire into their causes. In principle, an amoral anthropologist from another planet might do the same, studying these aspects of human thought and behavior much as we might study forms of communication and behavior in an ant colony, looking for the causal factors that give rise to them.

Scientific approaches to the study of morality adopt an external perspective in this sense, taking as their subject matter a set of empirical phenomena associated with moral judgments

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

and employing scientific methodologies to provide causal explanations of them. It is, after all, an empirical fact that human beings make judgments of this kind, employing moral concepts to condemn or praise actions or agents and to call for appropriate responses, with observable influences on social behavior. From a scientific point of view, then, we might begin by asking how human beings first came to possess the *capacity* to make such judgments (Joyce 2006; Kitcher 2006; 2011; Sinnott-Armstrong 2008a; Machery and Mallon 2010). Is it just a byproduct of our general capacity for language and thought, contingently developed in cultural contexts, as we might develop a capacity for judgments about fashion design by just extending the application of our more general psychological capacities? Or is the capacity for moral judgment and motivation instead an adaptation that evolved through natural selection much earlier in human history, because of the fitness advantages it conferred on our Pleistocene ancestors?

If Darwinian evolutionary forces were indeed involved, might they be responsible not only for the generic capacity and disposition to make and to respond to moral judgments, but also for dispositions to make certain *particular kinds* of moral judgment? For example, human beings may have been evolutionarily shaped to make (and to be motivated by) judgments condemning incest or in-group aggression or cheating, supported by evolved dispositions to experience certain particular kinds of emotional reaction involving disgust, guilt, blame, resentment, and sympathy, insofar as all of this was fitness-enhancing in ancestral environments. Those who speak of "the evolution of morality" typically have some or all of these sorts of claim in mind, where "morality" just refers to such capacities and dispositions of feeling, judgment, and behavior, as well as the moral codes and practices that have emerged in societies as a result of their exercise (Joyce 2006; Sinnott-Armstrong 2008a; Machery and Mallon 2010; Kitcher 2011). Anthropologists, sociologists, and historians might also conduct investigations into the particular types of moral judgment made over time in various cultures, looking at when and where certain concepts and applications first emerged and the patterns of thought, emotion, and behavior that grew up around them; and psychologists and neurophysiologists perform studies on people engaged in moral thought or experiencing moral feeling, investigating the factors that influence feeling and judgment, or the regions of the brain that are most active when subjects have thoughts such as M, or the neurochemicals important to the development of moral emotions, and so on (Greene 2008; Sinnott-Armstrong 2008b,c; Merritt, Doris, and Harman 2010; Churchland 2012).

All of these inquiries again target various *empirical phenomena* associated with talk of "morality," under the assumption that answers to them may appropriately be sought, as with any other empirical phenomena, through scientific inquiry. This is an unsurprising approach to the study of morality on the part of those who are primarily interested in providing causal explanations of empirical phenomena and see this sphere of human life as an interesting part of the world to be explored in this way. This might be described as a *scientific naturalist* approach to morality (not to be confused with philosophical and ethical naturalism, discussed later), and it promises to yield significant insights into a variety of interesting empirical phenomena.

The questions raised about morality from the scientific naturalist perspective, however, do not exhaust the questions most of us find it important to ask in this domain. Return to M and consider it this time from the *internal* perspective of a committed moral agent, as a judgment you yourself make. From this perspective, a different set of questions emerge – questions that are *philosophical* in nature, rather than scientific. When we make a judgment such as M, we again typically have *reasons* for it – considerations we cite as *justification* for

believing M to be *true*, and often these are the very things we take to *make* M true. In this case, for example, we will cite certain features of human trafficking – such as its grossly exploitative nature, its violation of people's fundamental rights and dignity, and its causing immense misery and psychological damage – as *wrong-making* features, and hence features that seem to make a judgment of wrongness, such as M, true. This then raises a variety of philosophical questions about these things: How exactly should we understand the meaning of moral judgments, such as M, and can they literally be true or false? If so, and if some can be true, what are their truth conditions? That is, what grounds *moral facts*, if they exist, and what is the nature of *moral properties* such as wrongness? What makes something a *good reason* for believing a moral proposition such as M, and what makes a feature of action count as genuinely *wrong-making*? Or are there no such normative reasons, properties, and facts in the moral domain after all?

These are not the sorts of question that occupy scientists looking at moral judgment simply as something that happens in the world (in the minds of human beings). Their concern is with the *empirical fact* of *someone's making* a judgment such as M, or in a group's incorporating it into a moral code, not with the *content* of M itself and with whether it might actually be *true*, and why. The latter is not a matter to be explored and settled scientifically, but a philosophical matter having to do with the meaning of moral judgments and their truth conditions (if any), to be explored through philosophical investigation and argument.

We seem, then, to have two very different sets of questions and methodologies concerning the study of morality, one scientific and the other philosophical. This naturally raises questions about how these two clusters of issues and forms of inquiry are related. Are these independent and complementary pursuits, each illuminating something important related to the notion of "morality" and staying out of the other's way? Or are they sometimes in tension with each other, fighting over some of the same terrain? Might the answers to at least some questions in one field actually depend on the answers to questions in the other? I will be suggesting that while there is a considerable degree of independence in these pursuits, there are also certain areas where they are in tension, and there is no simple resolution to that tension – certainly not one that can be settled by science itself. This is because the quality of the proposed scientific explanations in the disputed part of the terrain actually depends on philosophical issues that remain highly controversial. This ultimately casts doubt on whether good science, such as evolutionary biology, can be called upon to do the *debunking* work some philosophers have taken it to do in the moral domain. There may in fact be important limits to how far a *scientific naturalist* approach to morality can be pursued without begging important philosophical questions, and sensitivity to this issue should make us cautious about granting some of the explanatory claims debunkers purport to derive from the sciences, which lie at the heart of arguments purporting to debunk *moral realism* (roughly the claim that there exist real, knowable, objective moral truths; that is, moral truths that are not merely functions of our attitudes and that can be known by us).

Dividing the Terrain

Some scientific questions about morality can clearly be pursued independently of the concerns that animate moral philosophers. We don't require answers to any of these philosophical questions about judgments such as M in order to study the brain activity

exhibited by people making such judgments, or the effects that various environmental manipulations might have on the judgments they make – though philosophical input can be helpful in delineating moral versus nonmoral judgment, or in distinguishing different kinds of moral judgment in the experimental design. Similarly, the historical questions about the emergence of the human psychological capacities involved in moral judgment can appropriately be pursued without worrying about moral philosophers' concerns with the existence and nature of moral truth. Whatever we might think about the latter, it is clear that human beings have come to possess the capacities in question, and it is a broadly biological question whether they are adaptations shaped through natural selection due to their enhancement of biological fitness in ancestral environments, or whether they are instead side-effects of more general psychological adaptations that evolved for other reasons. It is likewise a broadly biological and psychological question whether certain psychological dispositions to make certain kinds of moral judgment (or underlying emotional dispositions influencing such judgment) can be traced back at least in part to psychological shaping by Darwinian evolutionary forces. Much of the work in this area will be highly speculative, but it is still something scientists can take up without consulting moral philosophers.

It is equally clear that certain philosophical questions about morality can be pursued independently of the concerns that animate scientists. In taking up first-order moral questions about how we should live and why, for example, it makes no difference how we came to have the capacity to think about such things – whether it evolved as an adaptation or is a spin-off from more general capacities. Nor will answers to such normative questions be settled by evolutionary claims about how and why we may have come to have certain behavioral, emotional, or cognitive dispositions. Suppose it were somehow established that a disposition to philander is an adaptation in male human beings, which evolved for the same Darwinian reasons that such dispositions have evolved in other species; perhaps related emotional dispositions evolved as well, leading to a tendency to form asymmetric attitudes toward male and female sexual infidelity, and hence to make asymmetric normative judgments about them that have historically been embodied as double standards in moral codes. Even if all of that turned out to be true, it would obviously not settle the relevant *normative* questions for us today, showing that such behavior is actually *good* or *right*, or that people have *good reason* to act on that disposition. That is a matter for us to settle through informed moral reflection on what makes for a good life, consistent with the full range of values bound up with our best understanding of a life that is both flourishing and worthy of us as rational beings who can strive for the good and not merely for satisfying evolutionarily given imperatives.

Of course, informed moral reflection will always take into account relevant facts of life, and the sciences have much to contribute here, thus playing an indirect role in informing moral reflection. Scientific work on animal cognition, for example, may rightly inform our moral thinking about the ethical treatment of non-human animals, and scientific work on race can undermine empirical claims that have been used to support racist ethical views. Similarly, psychological studies can illuminate facts about human limitations and weaknesses, which might inform our thinking about moral virtues and how best to cultivate them, or how to organize personal and social life to avoid characteristic pitfalls of the sort highlighted in work on situationism (Sabini and Silver 2005; Merritt, Doris, and Harman 2010). The point is just that the scientific work cannot itself settle the relevant *normative* questions for us, which always go beyond the facts about what we are like or how we came to be that way. Some of the scientific issues (such as evolutionary speculation on the origins

of some of our moral capacities and dispositions) are mostly irrelevant to these questions, while others are relevant, but only indirectly insofar as they illuminate the relevant factual background.

Other philosophical questions are not normative in nature but equally go beyond the scientific and are not settled by the answers to the scientific questions. Consider the question of meaning: Are moral judgments *cognitive*, consisting in beliefs with propositional contents that can be straightforwardly true or false, or are they instead *noncognitive*, consisting only in conative states involving desires, attitudes, and commitments? This is an issue philosophers can plausibly take up independently of questions about evolution, for example, because either answer is equally consistent with the competing evolutionary hypotheses. Even if our capacity for moral judgment evolved as an adaptation, for example, that does not suggest that what evolved was a capacity for judgment with a noncognitivist semantics, as some have claimed (Kitcher 2006). All it requires is that the judgment in question was capable of affecting motivation and behavior in fitness-enhancing ways, and that is equally possible for cognitive judgment, even on a Humean view, given the compresence of related desires and emotions (Joyce 2006).

Where things get much more difficult, however, is when we turn to attempts to explain the actual moral judgments we ourselves make today, which brings us back to M and the two very different perspectives one can take on it. *What is the best explanation of why we hold the moral beliefs we do* – for example, why you believe M? The problem is that, unlike with other parts of the terrain, where moral philosophers defer to scientists (e.g., regarding the origins of the capacities we use in making moral judgments) or scientists defer to moral philosophers (e.g., regarding the meaning and possible truth or falsity of moral judgments), both scientists and moral philosophers have a stake in this issue, each finding it within their own purview to offer explanations of why we believe what we do in the moral domain. These explanations, however, can be very different in nature, setting up a potential clash that is central to current philosophical debate over the relevance of evolutionary biology to the possibility of our knowledge of objective moral truths (such as we may take M to be). We turn now, then, to these explanatory issues and how some have appealed to scientific explanatory claims in this domain to try to debunk the moral realist's claim that we can and do possess such moral knowledge.

Explaining Beliefs

From a scientific naturalist perspective, the explanation of people's having the moral beliefs they have is approached using the same tools employed in other explanations of empirical phenomena: we use scientific methodologies to identify causal factors that have brought about and shaped the phenomenon in question. The question will be: What *caused* these beliefs to be formed in the people in question? (Just as we asked earlier what *caused* the capacity for moral judgment to emerge in humans.) One might, for example, look for Freudian or Nietzschean psychological causes, Marxist socioeconomic causes, or various other historical and cultural influences. And more fundamentally, one might seek deeper evolutionary causes that shaped human moral psychology in the distant past, giving us various dispositions of desire, emotion, attitude, and cognition (thereby also conditioning subsequent cultural developments), due to Darwinian selection pressures operating on Pleistocene humans. This sort of methodology will seem appropriate insofar as such inquiry

is taken to be continuous with the investigation of other empirical phenomena associated with human life, such as our evolved craving for fatty foods. Indeed, many will assume, at least tacitly, that this is the *only* respectable way to approach the explanation of our moral beliefs. What else could there be to consider other than the *scientifically accessible causes* of those beliefs – the kind that, again, an amoral anthropologist from another planet ought equally to be able to grasp and apply in explaining the phenomena, just as he could in explaining basic food cravings? Call this assumption *scientific naturalistic exclusivity*.

The problem, however, is that this exclusivity assumption overlooks a crucial alternative, which will be apparent to anyone who thinks about a proposition such as M not only from an *external* perspective, wearing the scientist's hat, but also from an *internal* perspective, wearing first the hat of a moral agent and then the philosopher's hat in reflecting on the issues. Recall that when you make a judgment such as M, this is not merely something that is caused to occur in you, as a rash might be caused to occur on your skin after contact with poison oak: it is instead something for which *you have reasons*, just as in other cases of belief, such as mathematical, scientific, or philosophical beliefs. In all these cases, the default question to ask in seeking to explain your belief – the question we pose when we ask *why* you believe the proposition in question – is not merely *what caused* a belief with that propositional content to occur in you, but rather: What are *your reasons* for the belief (i.e., for believing that content to be *true*)? That is, we seek (at least initially) your *justification* for your belief: the considerations you took to show that its content is true, thus explaining why you believe it. This is how it is appropriate to approach the explanation of belief, at least in those cases where people believe things *for reasons*, in the sense that their beliefs are based on considerations that they judge to be *good* reasons for believing the contents in question to be true.

Suppose you hold the mathematical belief that the square root of 2 is irrational: it cannot be expressed as a fraction of whole numbers. How should we set about explaining this belief of yours? It would be insulting to you to start out by asking merely what caused you to have the belief, casting about for causal factors having nothing to do with the truth of the content of the belief. That might be appropriate in a case where you have no mathematical understanding and believe what you do merely because, for example, you have been subject to a hypnotist who arbitrarily caused you to have this belief, just as he caused your friend to believe she is a chicken. But that is not how things ordinarily are. Suppose, then, that we are instead dealing with a typical case involving mathematical understanding, where you arrive at the belief after studying some number theory and working through the classic proof. In that case, the appropriate question to ask, and what we mean when we ask *why* you believe that the square root of 2 is irrational, is: What are your *reasons* for thinking that this propositional content is *true*? And your answer will appeal to the proof in question, which amounts to a set of considerations you take to *show why it is true* that the square root of two cannot be expressed as a fraction of whole numbers, thus *justifying* your belief. Where there is this sort of mathematical understanding, we *explain* your belief via *justification* from your perspective as an understanding mind, not merely via *causation from an external perspective* that is blind to issues of truth and justification (as on the model of the rash or post-hypnotic suggestion).

The same may be said about your various scientific and philosophical beliefs as well, such as your biological belief in common ancestry or your philosophical belief, say, that mental states are not type-identical to brain states: the place to start the explanatory task is with your reasons for believing these things, such as the considerations you cite as evidence of common ancestry (e.g., the fossil record or comparative genomics) or metaphysical

arguments concerning the multiple realizability of mental states. Once we understand your reasons, we can then ask *why you took* those considerations to be *good* reasons for believing the content in question, and once this is clear we will have satisfactorily explained your belief. Sticking with the mathematical example, we can ask: Why did you take the proof to constitute a good reason to believe the conclusion (that 2 is irrational), and hence come to believe it? And if you're mathematically competent and are thinking clearly, the answer will presumably be that the proof is *sound*, and therefore *is* a good, justifying reason to believe the conclusion, and you *recognized* this to be the case. This is how we explain your belief, rather than by appeal, for example, to evolutionary, sociological, or psychological causal factors as such. Again, it would be insulting to your mathematical intelligence and understanding to propose that what explains your belief is just some causal factor operating independently of the truth of the content of the belief. And notice that this sort of factive explanation, unlike the external "merely causal" explanations, involves *engagement with mathematics*, such as talk of the soundness of the proof, or of what is a good mathematical reason for believing a certain proposition. This is not the sort of explanation that would be available to a mathematically blind anthropologist or psychologist (to pick up the parallel with the amoral anthropologist from before).

Even in cases where you make a mistake and come to believe a false mathematical proposition based on a faulty proof, we still start the search for explanation by looking to your reasons for the belief. The complication is just that we'll need a different explanation for why you took this proof to be a good reason for believing the proposition in question: we can't say that it was because it *is* a good reason and you recognized it to be such, since in fact it is not. So, in such a case of error, we'll need some further account of why you *mistakenly thought* it was sound when it wasn't, and that explanation will appeal to various causal factors. Notice, however, that our overall approach still involves a thorough engagement, in the explanatory endeavor, with issues of truth and justification that are central to the internal perspective of the believer. We do not generally approach the explanation of *beliefs* simply and directly from the external perspective of scientifically accessible causes, as we do with the explanation of other empirical phenomena.

This is not to deny that we can sometimes justifiably adopt a largely external perspective in explaining beliefs. In some cases, there might be good evidence that a belief is *not* held *for reasons* in the sense used here, but is merely caused by extraneous factors having nothing to do with the truth of the content of the belief, and that any reasons offered are merely *post hoc rationalizations* rather than part of a true explanation of why the belief is held (Haidt 2001). This is how it would be in the case of post-hypnotic suggestion. It could also be the case if someone were caused to believe that homosexuality was wrong merely due to an unconscious gut reaction of disgust when contemplating it, and then rationalized his belief by casting about for reasons, such as the thought that it was "unnatural." Alternatively, even in some cases where beliefs are indeed held for the reasons the agent gives in explaining why he holds them, it may nonetheless be most plausible that the best explanation of why he *takes* those considerations to be *good* reasons for believing the contents in question to be true is one that has nothing to do with actual truth and justification: they may in fact be bad reasons, and he takes them to be good reasons only because of certain extraneous causal influences coming from his psychology, culture, or history. This is how it would be if someone genuinely believed that homosexuality was wrong on the grounds that (as he thought) it was "unnatural," but the only explanation for why he believed this was one that appealed to causal factors such as contingent elements of his cultural heritage or upbringing, or

perhaps even evolutionary influences on human cognition about such things. Here scientific explanations from a largely external perspective will have an important role to play.

The crucial point, however, is that the move to explanation by appeal to mere causes in the sense used here is warranted only insofar as we *already* have good *independent grounds* for rejecting the sort of factive explanation illustrated earlier, where the agent believes what she does because she correctly recognizes good reasons for believing the content in question to be true. We don't *begin* by looking for mere causes of beliefs without consideration of issues of truth and justification from the internal perspective of the believer. Instead, we start with that perspective, and only if we find compelling evidence that the factive model fails to apply do we turn to one of the two alternative models just described; in the more extreme case, we then propose that the belief is merely caused by extraneous factors and not based on reasons at all (often contrary to the agent's own assertion that they are); in the less extreme case, we grant that the belief is explained by the agent's believing certain things to be good reasons for taking its content to be true, but we explain away the agent's holding *that* belief about reasons simply by appeal to extraneous causal factors, rather than by appeal to her recognition of good reasons as such.

Explaining Our Moral Beliefs

Let us return, then, to moral beliefs such as M, and to the question of how best to approach explaining them. We have seen that, since we at least typically have reasons for such beliefs, as with the other sorts of belief discussed already, the appropriate starting point for explaining them is likewise to look to the reasons we give as purported justifications for them. Instead of starting by asking what evolutionary or cultural factors *caused them to occur* in people, we begin by asking for the *reasons* people give for taking them to be *true*. So if we ask why you believe M, your explanation might be: because human trafficking is exploitative, causes misery and harm, and is a violation of human dignity – considerations you take to be *wrong-making features* of actions and hence *good reasons* for believing human trafficking to be wrong and for believing M to be true. The next question, then, is why you believe these normative claims about those features of human trafficking (i.e., that they are wrong-making and are good reasons for believing M). This is analogous to asking why you believe that the considerations in the mathematical proof are good reasons for believing the proposition that 2 is an irrational number, or that the considerations having to do with the fossil record are good reasons for believing in common ancestry, or that the considerations given in certain metaphysical arguments are good reasons for rejecting the type-identity thesis in the philosophy of mind.

In those nonmoral parallel cases, it is very plausible to suppose that the answer goes something like this: you believe that the considerations in question are good reasons for believing the relevant proposition because they *are*, and being competent in mathematics, biology, and philosophy, you have recognized that fact (though the philosophical case will naturally be more controversial; other examples are easy enough to generate if one disagrees in this instance). We thus explain your holding these beliefs by appealing to your appreciation of good reasons for believing them to be true, and while this story will naturally involve various causes along the way, the explanation *goes through this internal appreciation of justification by the believing mind*, rather than appealing simply to external causal factors as if the belief were merely another occurrence in you, such as your coming to have a certain accent. The crucial question, then, is whether the explanation of *moral* beliefs at least sometimes fits this model.

The *scientific naturalist exclusivity* assumption mentioned earlier says that it does not: even if we cite as part of the explanation of someone's believing M the further beliefs she cites as her reasons, we can proceed to explain *her taking these considerations to be good reasons* for believing M by looking straightway to the kinds of causal factor cited in scientific explanations, without pausing to consider whether they might actually *be* good reasons and she might be correctly *recognizing* that normative fact about them. This again appears to be at least a tacit assumption in much scientific writing about "the explanation of morality," when it comes to explaining moral beliefs and the behavior that stems from them; for this is exactly how the discussion typically proceeds, with direct moves to evolutionary, sociological, or psychological causal factors to explain such beliefs and behavior, without troubling to worry about the philosophical possibility that at least some moral beliefs are instead properly explained on the justification-involving model we apply to other beliefs. No one approaches the explanation of those other beliefs by automatically looking just to extraneous causal factors, which would again come across as an insult to your mathematical, biological or philosophical intelligence and understanding. But this is in effect the approach commonly taken for granted by those who approach "the explanation of morality" exclusively from the external, scientific perspective.

Now, this approach would be unproblematic if we could just assume the truth of *moral nihilism*. If we could just assume that there are no moral truths to grasp, and no normative facts to appreciate about certain properties being wrong-making or about certain considerations being good reasons for believing certain moral propositions to be true, then we could dismiss the possibility of explaining someone's moral belief by appeal to these things, on the model employed in explaining other beliefs. We could simply explain someone's holding a certain moral belief by appeal to various other beliefs that are in turn explained simply in terms of causal factors having nothing to do with the truth of the contents of the beliefs. It would be, in the moral sphere, analogous to explaining all mathematical or scientific or philosophical beliefs by appeal to hypnotists or cultural contingencies or evolutionary influences.

The problem, of course, is that the assumption of moral nihilism is an extremely controversial *philosophical* assumption: it is not something established by the sciences, and it is not something widely accepted in moral philosophy, certainly not as anything like a default (which isn't to deny that it's a live possibility). This means that the scientific naturalist exclusivity assumption therefore begs a fundamental and very large philosophical question from the start. If we are to avoid just begging that question, we cannot assume that the external, scientific approach applies to the explanation of our moral beliefs across the board in the way it plausibly applies to the explanation of how we came to possess our basic mental capacities. Whether it does, and if so to what extent, depends on the answer to central philosophical questions that remain deeply controversial.

If moral realism, as opposed to nihilism, turns out to be true, it may well be that the proper explanation of at least some of our moral beliefs is not one appealing to mere evolutionary or cultural causes, but one appealing to the *truth* of M and our *appreciation* of that truth through a *proper recognition of the good reasons there are for believing it*, just as with many other true beliefs. For example, the proper explanation of your believing M may be not merely that you believe it because of causal factors in your upbringing (just as someone who believes that M is false does so merely because of distorting causal factors in *his* upbringing), but rather that – thanks in part to a *good* upbringing – you have come to recognize that human trafficking *is* indeed wrong (i.e., M is indeed true) by coming to

correctly appreciate that such exploitation and causing of misery indeed *make* an act wrong. This is not the sort of explanation that can be captured from the scientific perspective. As with the parallel explanation of your mathematical beliefs, where you believe what is true for good reasons and with appropriate understanding, this explanation requires engagement with issues of truth and justification internal to moral inquiry.

Note that this moral realist picture, which implies that there are limitations to the *scientific naturalist* explanation of our moral beliefs, needn't itself be *non*-naturalistic: it is, in fact, compatible with *philosophical naturalism* and *moral naturalism*. The philosophical naturalist holds that the only properties and facts in the world are natural ones, and when combined with moral realism, this yields *moral naturalism*: the view that there are real moral properties and facts, and that these are themselves fully natural ones, exhaustively constituted by properties and facts of the sort investigated by the sciences, even if they do not themselves show up on the scientific radar. Even moral naturalists can agree that there are limits to the *scientific* naturalist explanatory project when it comes to explaining our moral beliefs. As moral realists, they can still think that a proper explanation of at least some of our moral beliefs (where we get things right for the right reasons) takes the form of the explanation given for M, appealing to the believer's understanding of relevant reasons and grasping of truths, rather than the sort of explanation that might be given by an evolutionary psychologist or a sociologist. Of course, other realists will be led by philosophical considerations to posit *non*-natural moral properties and facts, and they will therefore reject philosophical and moral naturalism (see Chapter 28). But the relevant point here is that both naturalist and non-naturalist realists will have reason to be wary of sweeping claims made from a scientific perspective about "the explanation of our moral beliefs" across the board. On the realist view, some of the true ones are likely to be explained quite differently from the false ones, appealing in part to an understanding mind's proper appreciation of issues involving justification and truth. Finally, we turn to why this matters for metaethics.

Debunking Arguments

Moral realists who wish to preserve our claim to significant moral knowledge need to hold that our moral beliefs are at least reasonably accurate, representing the objective moral facts reasonably well, and nonaccidentally so. But are we justified in thinking this to be the case? Evolutionary debunkers raise doubts about this by pointing out that evolutionary processes would not have given us our capacity for moral judgment, or dispositions to exercise it in certain ways, in order to help us accurately track moral truths as such (Joyce 2006; Street 2006). As with any other adaptation, what would have mattered to the shaping of human moral belief-forming dispositions by natural selection over evolutionary time was simply that their exercise led to behaviors that enabled Pleistocene humans to increase their genetic contribution to the next generation through enhanced reproductive success. Even if there are real moral truths such as M, it is not accurate perception of such truths and behaving in accordance with them as such that would have been rewarded by natural selection, but simply *whatever* moral judgments (true *or* false) and behaviors (good *or* bad) happened to lead ancestral humans to leave more offspring.

The situation here is similar to that with dispositions for religious belief: a disposition to believe in an omniscient god who will punish rule-breaking even when no humans are

looking would have been rewarded by natural selection, even if the belief is completely false, as long as it was fitness-enhancing, as by increasing conformity to social rules in ancestral humans and thus increasing group stability and enhancing cooperation, leading to greater reproductive success. Similarly, dispositions to believe that philandering or killing stepchildren is morally acceptable would have been rewarded no less than dispositions to care for one's own children, as long as those dispositions were fitness-enhancing in ancestral environments, leading to differential reproductive success. In contrast with dispositions involving perceptual belief, which were fitness-enhancing (when they were) by virtue of yielding beliefs that *accurately represented the facts* in question, enabling ancestral humans to navigate their environment more effectively, dispositions involving religious or moral belief were fitness-enhancing (when they were) *not* because they yielded beliefs that accurately represented theological or moral facts, but simply because they yielded beliefs that led to behaviors that ultimately increased reproductive output (Griffiths and Wilkins forthcoming). And then the claim is that once we come to appreciate this fact, we lose any justified confidence that evolution would have given us reliable moral belief-forming capacities and dispositions, in the sense of reliably tracking the moral truths posited by realists.

Now, in itself, this is not yet very problematic for the realist. The natural response is simply to grant the point but argue that we don't *need* evolution to have given us reliable moral belief-forming capacities and dispositions in order to have them now – any more than we need natural selection to have given us reliable belief-forming dispositions in the domain of algebraic topology, or quantum electrodynamics, or metaphysics in order to have them now. What we need in all these cases is only for evolution to have given us the raw psychological materials necessary for *us* to *develop* reliable belief-forming dispositions ourselves, through appropriate cultivation and training in cultural contexts involving relevant traditions of inquiry and practice. Truths about abstract mathematics or physics or philosophy had no more role in shaping the cognitive capacities and dispositions of Pleistocene humans than moral truths did, yet we have been able to develop those capacities and dispositions in ways that make for reasonably reliable, truth-tracking exercises of them today, exhibiting significant *autonomy* from the particular evolutionary influences that gave us the raw materials to begin with. (This is clearly so in the first two cases, and we at least commonly proceed as if it's also true in philosophy, rather than declaring the field hopeless from the start simply because evolution itself didn't make us reliable philosophers). Moral realists may therefore hold that we have plausibly done the same in the domain of moral inquiry, even if the degree of reliability is rather less than in some other domains (FitzPatrick 2014; 2015).

In order to raise more serious problems for realists, debunkers have to go further. One option would be to try to block this move, arguing that for some reason we cannot have developed relevantly autonomous and reliable moral belief-forming dispositions in this way, as we have with respect to so many other exercises of human intelligence – including the very kind of philosophical thinking the debunkers themselves are relying upon. Another option would be to make stronger claims about the actual explanation of our moral beliefs themselves, purporting to undermine our justification for them (at least under realist assumptions about what would be required for their truth). I will focus on the second strategy, which is the most common move employed by evolutionary debunkers.

The crucial question, then, is: What is the *best explanation* of our moral beliefs? A tempting thought for those with debunking ambitions is to adopt a *scientific naturalist* perspective

on this question, citing the virtue of *parsimony* (which has been so important to scientific inquiry) and the ready availability of scientific causal stories about the shaping of the contents of our moral beliefs. These familiar stories appeal to human psychology and the contingencies of cultural histories, and more recently to underlying evolutionary influences on human cognition, emotion, and behavior. They are understandably attractive insofar as they are rooted in the sciences and are parsimonious, appealing only to familiar scientific causal factors: there is no appeal to such apparently exotic items as moral truths or facts. And they are purported to explain our beliefs fully, providing a "complete non-moral genealogy" of our beliefs, leaving out nothing of relevant explanatory interest (Joyce 2006).

For those who place a premium on parsimony in this context no less than in others, such explanations thus seem like clearly the "best" ones. And in explaining our moral beliefs without recourse to moral properties or truths or facts, they may seem to undermine any justification we might have had for believing in moral properties or truths or facts – just as fully explaining all beliefs about witches in terms of natural causes, with no appeal to truths or facts about witches, seems to undermine any justification we might have had for believing in truths or facts about witches. Or, at least, this is so assuming, in both cases, that the properties or truths or facts in question cannot simply be *reduced* to the sorts of factor that *are* involved in the explanations, or to other properties or truths or facts we have independent reason to posit (Joyce 2006; 2013). Moreover, if all our moral beliefs are fully explained in terms of "morally blind" causal influences operating insensitively to any moral properties or facts as such – starting with evolutionary influences following principles of natural selection, and continuing with equally truth-blind causal influences from cultural inputs – then there is no reason to suppose that they are even approximately accurate (in the sense of tracking the moral truths posited by realists, even if they exist). Indeed, it would be a sheer *coincidence* if they were, and we have no justification for believing that any such coincidence has occurred (Street 2006; 2008). Once we realize this, then, *our justification for our moral beliefs themselves* would plausibly be *defeated* (again, under the realist assumption that for moral beliefs to be true is for them accurately to represent objective moral facts) – much as if we discovered that they were all merely caused by a hypnotist playing games with us. Moreover, even if many of our moral beliefs turned out to be true, the element of *accidentalness* involved would seem to undermine any serious claim to moral knowledge and understanding. We would be stuck with moral skepticism.

This basically captures the core themes of epistemic debunking arguments, though there are many subtleties and differences to be brought out in various versions, which are set aside for brevity here (for detailed treatment of such arguments, see, among many others, Shafer-Landau 2012; Berker 2014; FitzPatrick 2008; 2014; 2015).

Response to the Debunking Arguments

There are various possible responses to such debunking arguments. For present purposes, however, there is one that stands out as most salient given what has been said already. The crucial point is that the question of what the "best" explanation is for our moral beliefs is *itself* every bit as fraught as the questions that gave rise to the larger metaethical debate, about the truth or falsity of moral realism, in the first place. Naturally, those who have *already* rejected all forms of moral realism for other reasons will be strongly attracted to a scientific naturalist approach to explaining our moral beliefs. What else could explain

people's moral beliefs than the various causal factors investigated by the relevant sciences? The situation would be similar to that faced by the realist when it comes to explaining moral beliefs she takes to be false – such as beliefs steeped in superstition, racism, rigid conceptions of gender roles, the desire to maintain power and privilege, and so forth – except that it would be taken to apply across the board, to all our moral beliefs, however reflective and well-informed by relevant facts. By contrast, however, those of us who have *not already rejected* moral realism cannot be expected to find such an approach to the explanation of our moral beliefs to be the "best" one. And this appears to lead to an impasse on this issue, which obviously cannot itself be resolved by appeal to the sciences.

Debunking arguments are intended to show that there are no knowable, objective moral truths (or alternatively, that we are unjustified in believing in the existence of moral truths): even if there were objective moral truths, such as M, perhaps, *we* could not *know* them, because our justification for our moral beliefs is defeated by our recognition of the points in the debunking argument (or alternatively, our justification for our belief in the existence of moral truths is defeated by our recognition of such points). But unless we've *already* rejected the idea of knowable, objective moral truths, we aren't likely to grant the debunker's claims about best explanation to begin with. And, at the outset of the argument, it is supposed to be an open question whether there are such truths. So suppose we begin by thinking that there are, and that we know many of them, such as M. Then we will likely think that while scientific naturalist *stories* can always be told about why we believe what we do, those stories will often be lacking and misleading, because the *correct* explanation for why believe some moral propositions, such as M, is the factive one mentioned earlier, rather than an external one in terms of causal factors having nothing to do with the truth of M. That is, we will think that we believe some propositions, such as M, because we've *recognized* their *truth* through *recognizing good reasons* for believing them to be true, just as in the mathematical case involving competent mathematical judgment. And in that case, we will simply not grant the debunker's strong explanatory claims about the "real" explanation of our moral beliefs across the board, involving mere evolutionary causal factors and similarly morally blind cultural pushes and pulls. We will take at least many of our moral beliefs to have been relevantly *guided by the moral and justificatory facts*, rather than merely jostled by causes insensitive to moral properties and facts, and we will therefore reject the explanatory claims that were supposed to defeat our justification for our moral beliefs (or for our belief in moral truths), or to show that our beliefs are at best only accidentally true.

In response, debunkers will claim that they have parsimony on their side: it is more economical to appeal simply to scientifically accessible causal factors than to appeal to moral facts and good reasons (especially where these are construed non-naturalistically). But, as before, this appeal to the value of parsimony is *itself* as highly controversial in this context as the competing claims in the metaethical debate. While parsimony has played an important role in the sciences, part of what is at issue in the metaethical debate is precisely whether it should be assumed to play an equally important role here. Parsimony is a virtue only where the world is correspondingly spare; where it is not, insistence on parsimony will only lead to error or misrepresentation. For someone initially attracted to moral realism, parsimonious explanations of beliefs such as M from a scientific naturalist perspective – such as, "You just believe that because some cultural contingencies in the way you were raised caused a certain extension of an evolved sense of sympathy to victims of trafficking, causing you to have that belief about wrongness" – are as wrongheaded and insulting to the understanding mind as corresponding explanations of mathematical or philosophical

beliefs would be. Explanations offered from an external perspective citing "merely causal" factors and eschewing engagement with the issues of truth and justification internal to the domain of inquiry in question will seem to be missing something crucial to the proper explanation of the belief (even if some of the factors mentioned are part of the story insofar as they help explain how you came to appreciate the truths in question). If there is some theoretical cost in rejecting the scientists' greater parsimony, the moral realist may think there is a far greater cost in having to swallow the claim that our belief in the wrongness of human trafficking has nothing whatever to do with its actually being wrong.

Perhaps someone genuinely on the fence in the metaethical debate, with little attachment to moral realism to begin with, will be persuaded by the appeal to parsimony to reject realism, finding the scientific naturalist explanations more palatable in the end. But no one who instead starts out confident that she knows human trafficking to be objectively wrong, say, for the reasons she gives, can reasonably be expected to be moved by considerations of parsimony to accept merely psychological, sociological, or evolutionary explanations of her belief as being complete or exhaustive. Insofar as such explanations leave out any appeal to her proper recognition of the *moral truth* that trafficking is wrong, via a proper recognition of the reasons in question as *good* reasons for believing it to be wrong, they will seem far from complete or exhaustive; in fact, they will seem to be missing precisely what is most important. And this is how it plausibly is with many of us: our moral experience – from the perspective of committed moral agents – of apparently recognizing good reasons and grasping moral truths should preclude any rush to embrace such external, debunking explanations of our beliefs.

We could, of course, be deluded about our experience, just as we think those who hold false moral beliefs are, and if so, then the debunkers have plausible stories to tell about what is really going on. The debunking arguments may thus present a good challenge, adding evolutionary considerations to the familiar stories that, *if* true and exhaustive, would indeed shake our confidence in our moral beliefs and put the idea of moral truths on shaky epistemic ground (Joyce 2013). But it is one thing to be able to tell debunking stories about the exhaustive etiology of our moral beliefs and quite another for those stories to be correct. The stories by themselves do not debunk anything, but merely remind us that such debunking is a possibility we cannot ignore, and show us what it might look like. Their correctness – that is, the truth of the proposed debunking explanations, when taken to be exhaustive – would indeed cause problems for moral realists. But whether such stories are correct or not is no less controversial than the debate over moral realism in the first place.

References

Berker, S. (2014). Does Evolutionary Psychology Show that Normativity is Mind-Dependent? In *Moral Psychology and Human Agency: Essays on the New Science of Ethics*, edited by Justin D'Arms and Daniel Jacobson. Oxford: Oxford University Press.

Churchland, P. (2012). *Braintrust: What Neuroscience Tells Us About Morality*. Princeton: Princeton University Press.

FitzPatrick, W.J. (2008). Morality and Evolutionary Biology. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/archives/sum2012/entries/morality-biology/ (last accessed July 15, 2015).

FitzPatrick, W.J. (2014). Why there Is No Darwinian Dilemma for Ethical Realism. In *Challenges to Moral and Religious Belief: Disagreement and Evolution*, edited by Michael Bergmann and Patrick Kain. Oxford: Oxford University Press.

FitzPatrick, W.J. (2015). Debunking Evolutionary Debunking of Ethical Realism. *Philosophical Studies* 172(4): 883–904.

Greene, J. (2008). The Secret Joke of Kant's Soul. In *Moral Psychology,* vol. 3: *The Neuroscience of Morality*, edited by Walter Sinnott-Armstrong, pp. 35–80. Cambridge, MA: MIT Press.

Griffiths, P. and Wilkins, J. (Forthcoming). When Do Evolutionary Explanations of Belief Debunk Belief? In *Darwin in the 21st Century: Nature, Humanity and God*, edited by Philip Sloan. Notre Dame: Notre Dame University Press.

Haidt, J. (2001). The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108(4): 814–834.

Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

Joyce, R. (2013). Irrealism and the Genealogy of Morals. *Ratio* 26(4): 351–372.

Kitcher, P. (2006). Biology and Ethics. In *The Oxford Handbook of Ethical Theory*, edited by David Copp. Oxford: Oxford University Press.

Kitcher, P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.

Machery, E. and Mallon, R. (2010). Evolution of Morality. In *The Moral Psychology Handbook*, edited by John M. Doris, pp. 3–46. Oxford: Oxford University Press.

Merritt, M., Doris, J., and Harman, G. (2010). Character. In *The Moral Psychology Handbook*, edited by John M. Doris, pp. 355–401. Oxford: Oxford University Press.

Sabini, J. and Silver, M. (2005). Lack of Character? Situationism Critiqued. *Ethics* 115: 535–562.

Shafer-Landau, R. (2012). Evolutionary Debunking, Moral Realism and Moral Knowledge. *Journal of Ethics and Social Philosophy* 7(1): 1–37.

Sinnott-Armstrong, W. (ed.) (2008a). *Moral Psychology,* vol. 1: *The Evolution of Morality*. Cambridge, MA: MIT Press.

Sinnott-Armstrong, W. (ed.) (2008b). *Moral Psychology,* vol. 2: *The Cognitive Science of Morality*. Cambridge, MA: MIT Press.

Sinnott-Armstrong, W. (ed.) (2008c). *Moral Psychology,* vol. 3: *The Neuroscience of Morality*. Cambridge, MA: MIT Press.

Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127: 109–166.

Street, S. (2008). Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About. *Philosophical Issues* 18: 207–228.

28

What's to be Said for Moral Non-Naturalism?

TERENCE D. CUNEO

In the not too distant past, many philosophers summarily dismissed moral non-naturalism. Though these philosophers acknowledged that non-naturalism has an impressive pedigree, having been defended by figures such as Plato and G.E. Moore, they charged it with being unmotivated, ontologically exorbitant, even "spooky" (see Jackson 2012). Christine Korsgaard, for example, describes the view as one according to which there are moral facts "wafting by" that we somehow grasp (1996, 44). Allan Gibbard characterizes it as requiring of us special powers to "peer into a special realm of normative fact" (1990, 284). Though these images of non-natural facts as belonging to some ethereal domain continue to operate on the thinking of moral philosophers, they no longer exert the influence that they once did. Thanks to the efforts of philosophers such as Russ Shafer-Landau (2003), David Enoch (2011), and Derek Parfit (2011), non-naturalism is once again a view that many philosophers take very seriously.

My project in this chapter is to give you a feel for why an increasing number of philosophers take non-naturalism seriously, presenting a two-stage argument in its favor. The first part of the argument presses the point that moral naturalists must defend some form of reductionism according to which moral properties are "nothing over and above" natural or descriptive properties. Vindicating reductionism, I contend, is considerably more challenging than many have assumed. The second part of the argument develops the claim that once we specify what moral facts would have to be like, there is reason to hold that some moral facts are as non-naturalists believe.

At the end of this chapter, I will indicate why I believe that the argument I'll present is importantly incomplete. Before moving forward, however, I should flag two other respects in which the argument has limited ambitions. First, any full defense of non-naturalism would have to specify how non-naturalists can respond to important challenges to their view, of which there are several. While I believe that such a defense can be provided, I'll not attempt to offer it. Moreover, any defense of "robust" non-naturalism would have to explain why thinner versions of the view, which claim that while there are non-natural moral truths,

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

these truths have "no positive ontological implications," are unsatisfactory (see Parfit 2011, 478; see also Scanlon 2014). I'll not attempt to do that on this occasion. Instead, I'll assume that non-naturalism makes substantive and controversial ontological commitments, specifically concerning the nature of moral facts.

What is Moral Non-Naturalism?

Let us begin by identifying several points on which moral naturalists and non-naturalists agree. First, philosophers of both sorts maintain that there are moral facts, such as *that it is wrong to cheat others simply for gain*. Second, both (as I shall think of them) are moral realists, agreeing that some moral facts are objective, in the sense that they hold independently of what we value or what conventions hold. According to realists, the fact to which I just adverted does not exist because some of us are against cheating or because there are social conventions that prohibit cheating for gain. Finally, both naturalists and non-naturalists agree that moral facts do not float free of the natural realm but bear intimate relations to it; actions are wrong, these philosophers agree, because of the way things are at the natural or descriptive level.

Though naturalists and non-naturalists agree on this much, they disagree about the nature of moral facts. Moral naturalists hold that moral facts are nothing over and above natural ones. Moral non-naturalists demur, holding that some moral facts depend on but are not themselves natural facts. While many philosophers believe that this difference between naturalism and non-naturalism is deep and important, offering an informative specification of what distinguishes natural from non-natural facts has proven extraordinarily difficult. All the proposals in the literature are problematic.[1]

Given our topic, it would be helpful to have an informative account of what distinguishes natural from non-natural facts. While I would very much like to have such an account, I have none to offer. In its absence, I propose to employ a less-than-fully-informative but nonetheless expedient route, which is to assemble two lists. The first includes terms that putatively refer to entities that most parties to this debate would identify as being natural, perhaps because they are such as to play explanatory roles in the natural sciences or are reducible to such features. The second includes putatively referring terms that don't make it on to the first list. According to this strategy, if a term doesn't make it on to the first list, then it is a good candidate for being an expression that might refer to something non-natural.

Nothing about this approach, I should emphasize, implies that terms on either list refer. Nor does this approach imply that if these terms do refer, those on the first list refer to natural entities, while those on the second list refer to non-natural ones. After all, for all I say here, Berkeleyan idealism (which holds that everything that exists is nonphysical) might be true, in which case many of the terms on the first list would either not refer or would refer to non-natural entities. Or, alternatively, reductive physicalism might be true, in which case many of the terms on the second list would not refer or would refer to natural entities. The role of compiling these lists, then, is heuristic; it is simply to furnish promising candidates for entities that would be natural or non-natural entities (if there are any).

1 Copp (2007, ch. 1), Cuneo (2007), McPherson (2015), and Shafer-Landau (2003) offer different ways of understanding the distinction.

Let's consider what sorts of term these lists might include. A version of the first list should probably include terms that putatively refer to objects of various sorts, such as the planet Mars, quarks, and elephants. Likewise, it should probably include terms that refer to properties such as *having negative charge*, *being in pain*, and *having the atomic number 8*. And, finally, it should probably include terms that refer to concepts that correspond to these properties, such as HAVING NEGATIVE CHARGE, BEING IN PAIN, and HAVING THE ATOMIC NUMBER 8. The second list might include terms that are not included on the first, such as expressions that putatively refer to objects such as God. This list might also include terms that refer to properties such as *being sublime*, *being a promise*, or *counting in favor of*. And, finally, it should include terms that refer to concepts that correspond to these properties, such as BEING SUBLIME, BEING A PROMISE, and COUNTING IN FAVOR OF.[2]

Non-naturalists hold that terms that refer to moral concepts and moral properties belong on the second list. The rationale that non-naturalists offer for this claim is that moral concepts and properties are importantly different from the sorts of concept and property that belong on the first list. Think, for example, of moral concepts such as BEING WRONG and BEING UNJUST. These concepts are not such as to play any interesting explanatory roles in our present sciences, including physics, chemistry, and biology. Nor do we have any reason to believe that they would be fit to play any such role in a perfected or idealized version of these sciences, since we have no clear idea what these idealized sciences would be like. Nor, moreover, do these concepts appear to be reducible to concepts that are such as to play explanatory roles in the sciences, such as HAVING NEGATIVE CHARGE, HAVING THE ATOMIC NUMBER 8, or BEING FITNESS ENHANCING. As philosophers since G.E. Moore have emphasized, attempts to argue that a moral concept such as RIGHT is identical with a natural concept such as BEING SUCH AS TO ENHANCE EVOLUTIONARY FITNESS have come to grief, since there is nothing incoherent or inapt in thinking that an action enhances evolutionary fitness but is not right. (Selective cheating, for example, might enhance human evolutionary fitness but is, in the ordinary case, not right.)

The line of argument articulated in the last paragraph suggests that there is a more general point to be pressed against any attempt to identify moral with natural concepts. For, by all appearances, natural concepts have a certain function or job description, which is (roughly) to figure in theories that predict and explain occurrences in the natural world, such as happenings on the quantum level or the behavior of gasses. Moral concepts, however, appear to have a very different job description. Their primary function is not to figure in theories that predict and explain what happens in the natural world, but rather is *practical*: they are employed to help guide decisions and evaluate behavior. For example, by noting that you have a decisive moral reason to return a book that you've borrowed from me, you can determine how you ought to act. If you fail to do what you should, we can employ moral concepts such as WRONG or INCONSIDERATE to evaluate what you've done (or not done). On the basis of the application of these concepts, we can then determine whether to hold you accountable, admonish you, blame you, or the like for failing to return the book. When we do so, we are not trying to explain or predict your actions; we are evaluating them. We are thinking "ought thoughts."

There is, then, a very general reason to believe that natural and moral concepts are fundamentally different, as they appear to play different roles in our lives. Since this claim

2 Like Gibbard (2003; 2012), I hold that concepts can be non-natural. For a brief defense of this claim, see Cuneo and Shafer-Landau (2014).

is important to the following argument, it will be helpful to have a name for it. I'll call it *the explanatory/normative divide*. Having identified this claim, it is now important to highlight a distinction that was largely implicit in the preceding paragraphs – a distinction between *concepts*, on the one hand, and *properties* and *facts*, on the other.

To get a handle on this distinction, begin by noting that we often operate with a rough and ready distinction between thoughts and world – between what we think and what there is. Concepts belong to the "thought" side of this distinction, since (as I shall understand them) they are abstract and sharable ways of thinking about objects and properties. By employing concepts such as HORSE, NEIGHBOR, and SKITTISH, for example, I can form the thought THAT MY NEIGHBOR'S HORSE IS SKITTISH – this thought or proposition having as its constitutive components the concepts HORSE, NEIGHBOR, and BEING SKITTISH. Like many other thoughts, this thought purports to be about the world: in this case, my neighbor's horse. Once we see this, we can also see that concepts are not merely the subcomponents of propositions, they are also *referential devices*. They are that which enables thinkers to think about or refer to things such as horses and skittishness – if any such things there be.

Unlike concepts, properties belong to the "world" side of the thought–world distinction, since they are features that entities can have or exemplify. Suppose, to stay with our example, that my neighbor's horse actually is skittish. If it is, then this horse has the property of *being skittish*. When it does, we can say that it is a fact *that my neighbor's horse is skittish*, as many facts are simply situations or states of affairs in which things have properties. Understood in this way, facts often make propositions true. My thought THAT MY NEIGHBOR'S HORSE IS SKITTISH is made true by the fact *that my neighbor's horse is skittish*. If I also form the moral thought THAT MY NEIGHBOR HAS TREATED HIS SKITTISH HORSE WRONGLY, then this thought is true if and only if and because it is a fact *that my neighbor has treated his skittish horse wrongly*.

The reason that I am calling attention to this distinction between thought and world is this: often we think of the same thing by using different concepts. If you are baking a cake, you might have the thought that you need to turn up the heat in the oven that you're using. If, by contrast, you are writing a paper in physics, you might have the thought that scientists have failed to understand certain properties of mean kinetic energy. While in both cases you would be thinking different thoughts, you would nonetheless be thinking about the same thing – namely, mean kinetic energy – but in different ways. Otherwise put, while the concepts HEAT and MEAN MOLECULAR KINETIC ENERGY are different, they refer to the same thing. This observation opens the conceptual space for the following possibility, which is important in the present discussion.

The possibility is one according to which non-naturalists are correct to claim that moral concepts such as WRONG cannot be identified with natural or descriptive concepts but incorrect to hold that these concepts characterize non-natural properties or facts. This possibility holds not because moral concepts fail to refer at all, as error theorists believe. Rather, it is because it seems possible that moral concepts characterize only natural or descriptive properties or facts. To those familiar with debates in the philosophy of mind, the structure of this view should seem familiar, as it would be analogous to a view according to which concepts such as PAIN or TASTING CHOCOLATY are not identical with concepts such as BEING IN SUCH-AND-SUCH BRAIN STATE, but nonetheless refer (only) to brain states. We might, of course, be completely ignorant of the fact that moral and mental concepts work that way. However, successfully referring to something by the use of a concept needn't imply that those who employ such a concept have correct views about how that concept works or to what it refers.

The debate between non-naturalists and naturalists has various dimensions, and in our discussion so far we have only touched upon some of them. Even if we do not explore the various dimensions of this debate – such as what motivates philosophers to accept one or another of these positions – we are in a position to see that a key point of disagreement will be whether it's plausible to believe that moral concepts, which by all appearances are not natural, refer to natural properties or facts. Non-naturalists must offer reasons for believing that they do not and naturalists must furnish reasons why they do. Who should we believe? In the next section, I discuss two naturalist positions, each of which employs different strategies to contend that, while moral concepts are non-natural, moral properties and facts are natural. The upshot of the argument I present is that these strategies are deeply problematic.

Two Reductionist Strategies

Suppose you believe that everything that exists is part of the natural world. If you also believe that there are moral facts, as moral naturalists do, then you are committed to the claim that they too must be part of the natural world, since they could be nothing over and above natural facts. In claiming that moral features are nothing over and above the natural world, moral naturalists endeavor to reduce moral features to natural ones.

As it happens, naturalists have not agreed about what it is for something "not to be over and above" the natural realm. According to some naturalists, such as Frank Jackson, the "not over and above" relation is simply that of identity (see Jackson 1998). Modified somewhat, Jackson's argument for the thesis that moral features are identical with natural ones runs as follows.[3] Moral features such as *being wrong* do not float free of natural properties such as *causing pain*; rather, they supervene on them, as there cannot be a change in a thing's moral features without a change in its natural features. Now, suppose we were to bundle together all the natural features on which a given moral feature – say, wrongness – could supervene, creating a huge disjunctive natural property. Call this property N. N would be necessarily equivalent with wrongness, for, necessarily, something would be wrong if and only if it were N. But, says Jackson, we should accept the sparse theory of properties according to which properties that are metaphysically equivalent are identical. It follows from the sparse theory that wrongness is identical with N, which is a natural property. Since there is nothing unique about wrongness in this respect, it follows that every moral property is identical with a natural one.

When understood as a view about facts, Jackson's argument faces an especially pressing worry, which is its move from:

Fact A and fact B are metaphysically equivalent: necessarily, A exists if and only if B does.

to:

Fact A and fact B are identical.

3 I am omitting various details of Jackson's position, including the fact that he is an analytic naturalist, holding that moral and descriptive terms share the same semantic content.

One concern about this move is that it yields bizarre results. After all, if it were correct, it would follow that the fact *that there is such a person as God or it is false that there is such a person as God* is identical with the fact *that 2 + 3 = 5*, since they are metaphysically equivalent. However, these facts certainly do not seem to be identical; one concerns numbers, the other doesn't (see Plantinga 2010).

Another worry, perhaps more pressing than the first, is that properties and facts can bear different sorts of relation to one another, including that of *being more fundamental* than another or *being such as to determine* another, even when they are metaphysically equivalent. However, Jackson's view, which takes metaphysical equivalence to be sufficient for identity, is insensitive to determination relations that might hold when one property or fact is more fundamental than another.

To illustrate, consider the fact *that sugar is sweet*. This fact is metaphysically equivalent to the fact *that an omniscient agent would believe that sugar is sweet*. But, on the face of things, this provides little reason to believe that these facts are identical. To the contrary, it seems that the second fact holds in virtue of the first; an omniscient observer would believe that sugar is sweet because it is. When applied to the moral domain, we can raise the same concern. Suppose, for argument's sake, that the properties *being morally obligatory* and *being such as to maximize happiness* are metaphysically equivalent. One possibility is that the latter property determines the former; another possibility is that they are identical. Short of denying that there are any determination relations at all, however, the fact that these properties are metaphysically equivalent would not itself imply that they are identical (any more than the fact *that sugar is sweet* is metaphysically equivalent with the fact *that an omniscient being would believe that sugar is sweet* establishes that these facts are identical).

For those who believe that facts or properties can determine one another, then, there are good reasons to be uneasy with Jackson's reductionist proposal. There is, however, another variety of naturalism that rejects the idea that the "nothing over and above" relation is best characterized in terms of identity. Philosophers who espouse this view maintain that to reduce moral features to natural ones is to show that natural features *constitutively explain* moral ones. To get a feel for this view, consider an example from the non-moral realm: the property *being a triangle* is constituted by *having*, *three*, *angles*. Something is a triangle because it has these constituents, but these constituents are not what they are in virtue of constituting a triangle. If this is right, reducing a property F to a property G is not so much a matter of establishing that F and G are identical as establishing that the constituents of G constitutively explain F.

Mark Schroeder, who defends this position, maintains that any putative reduction of this (or indeed any) sort will be successful only when a pair of conditions is met (see Schroeder 2007). For one thing, any object or property that is a candidate for being reduced will have "fixed features": features that we have strong reason to believe must be true of it. Most believe, for example, that to be in a state of pain is to feel a certain way (call this feel being "ouchy"). A successful reduction of pain states to brain states, then, must imply that pain states are "ouchy," this being a fixed feature of pain states. Moreover, any successful reduction must leave enough of our ordinary background picture of the world intact. A reduction is not successful if it implies that pain states are "ouchy" but that "ouchiness" is not a feel, but simply a tendency to wince or to curse. This broadly behaviorist construal of ouchiness is unacceptable because it is a deeply entrenched part of our background picture of the world that we can experience feels of various sorts. Could any putative reduction of the moral to the natural satisfy these two conditions? Philosophers such as Schroeder believe so. If we

understand reduction in terms of constitutive explanations, these philosophers hold that fundamental normative features, such as *being a reason*, might well reduce to natural features, such as *being desired*.

Let's consider the constitutive explanation strategy in more detail, focusing on Schroeder's proposal, since it is the most developed. Schroeder's case for his view begins with a simple thought experiment: imagine that there is going to be a party tonight, at which there will be dancing. Imagine, further, that while Ronnie loves to dance, Bradley hates it. On the face of things, the fact that there will be dancing at the party is a reason for Ronnie to go, since he loves to dance. In contrast, the fact that there will be dancing at the party is a reason for Bradley not to go, since he abhors it. If this is right, we know that in some cases, whether an agent has a reason to act in some way is determined by features of his psychology; specifically, his desires. Might all of an agent's reasons be explained in this way? Schroeder suggests that the answer is yes: for any fact whatsoever, if that fact is a reason for an agent to act, then we can explain it in terms of that agent's desires.

Put more formally, Schroeder's proposal is what he calls:

> **Reason** For [a proposition] R to be a reason for X to do A is for there to be some [proposition] p such that X has a desire whose object is p, and the truth of R is part of what explains why X's doing A promotes p. (2007, 59)[4]

Stated more simply, **Reason** tells us that an agent's reasons to act are those propositions whose truth helps explain why acting in some way increases the likelihood of that agent's satisfying one or another of his desires.

Before critically engaging Schroeder's proposal, let me try to clarify it further. First, Schroeder's view is not that desires are reasons but that desires constitutively explain why certain features of the world – such as that there will be dancing at a party – are reasons. Thus understood, desires are "reason-explainers." Second, Schroeder has a very liberal understanding of the promotion relation: anything that raises the probability of an agent's satisfying a desire "promotes" the satisfaction of that desire (2007, 113). Third, an implication of this last point is that, for just about any action whatsoever, there will be a reason for an agent to perform it, since that agent will have *some* desire such that performing that action will raise the likelihood of his satisfying that desire. It does not follow from this last point, of course, that if you have a reason to perform some action, then you *should*, since reasons can have different weights. As stated, **Reason** tells us nothing about how much weight a particular reason might have for an agent (see Schroeder 2007, ch. 7).

It is easy to see that **Reason** is a very controversial proposal, since it is very inclusive. It implies, among other things, that you now have reason to chew on your car – the explanation being that since you desire to be healthy and chewing on your car will introduce healthful iron into your diet, the likelihood that this desire will be satisfied will be increased if you chew on your car. **Reason** also implies that, given that Ronnie has certain desires – such as not wanting to injure himself – he has reasons not to dance, and that, given some of Bradley's desires – such as wanting to be more physically fit – he has reasons to dance. This is odd, because the case of Ronnie and Bradley was supposed to be one that illustrates that

4 According to this formulation, reasons are propositions. Elsewhere, Schroeder speaks as if they are facts or events. While I believe that propositions, facts, and events belong to different ontological categories, when I sketch Schroeder's view, I'll not pay attention to these differences.

they have reasons to act differently. Given Schroeder's view, it does not illustrate this at all, since the fact that there is dancing at the party is, for each of them, both a reason to attend and a reason not to attend (see McPherson 2012).

Suppose, though, that we waive such initial concerns, assuming that **Reason** is correct. If it is, it will follow that an agent's having a reason to act in a certain way is reducible to natural features, namely, its being such as to raise the likelihood of that agent's satisfying a desire. The question is whether this proposal is satisfactory.

Let me offer two reasons for believing not. First, it is one thing to show that a particular kind of normative feature, namely, being a reason, is reducible to natural features; it is another thing to establish that *all* normative features are reducible to descriptive features, which is what a naturalist position must establish. Schroeder expresses optimism that if reasons are reducible to natural features, then normative features as such are reducible to natural ones, since we can constitutively explain all normative features in terms of reasons (2007, ch. 4). While I am not prepared to say that such a reduction is impossible, I see no reason for being optimistic about its prospects.

Here is one reason why: Suppose you have a moral right against me that I not attack you (and, hence, I have an obligation not to attack you). And suppose that I have the same right against you (and, hence, you have the same obligation toward me). A "reasons-first" approach of the sort that Schroeder favors must constitutively explain our having this right in terms of reasons. Moreover, on the assumption that we each have this right in virtue of some shared characteristic, it follows that, under Schroeder's position, there must be some desire such that each of us has it, and the likelihood of its being satisfied is increased by my not attacking you and your not attacking me. But it is (to put it mildly) highly contentious that there is such a desire. The desire to avoid bodily harm is not such a desire, for it is possible that I lack this desire; if that is the case, however, it doesn't follow that it is morally permissible for you to attack me.

The second concern about Schroeder's proposal is whether it satisfies its own strictures for being successfully reductive. Recall that, according to Schroeder, any putative reduction will be successful only when the following is true. First, any object or property that is a candidate for being reduced will have "fixed features": features that we have strong reason to believe must be true of it. Moreover, any successful reduction must leave enough of our ordinary background picture of the world intact. The attempt to constitutively explain triangularity in terms of *having*, *three*, *angles* seems to satisfy these criteria; it does no violence to our ordinary understanding of triangularity – to the contrary, it illuminates it. The attempt to constitutively explain the "ouchiness" of pain in terms of behavioral dispositions, in contrast, appears to fail these criteria, since pain appears to be not a behavioral disposition but a feeling. This attempt at reduction does great violence to our ordinary understanding of pain – such violence that an eliminativist about the mental (someone who denies that there are any mental events as we ordinarily think of them) could probably accept it. The question to ask is whether **Reason** more nearly resembles the mathematician's attempt to constitutively explain triangularity or the behaviorist's attempt to constitutively explain pain.

Reason, I believe, is closer to the behaviorist's attempt to constitutively explain pain, and for a very fundamental reason: **Reason** reduces reasons to an explanatory relation, since it tells us that a reason for you is simply what raises the probability that a desire of yours is satisfied (see Olson forthcoming). But normative relations such as *being such as to favor* and *justifying* are, by all appearances, not merely explanatory relations. After all, it seems possible for a fact – such as *that there will be dancing at the party tonight* – to help explain

another fact about me and my mental life but not to favor or justify my behaving in any way at all. For example, suppose the fact *that there will be dancing at the party tonight* raises the likelihood that some *belief* of mine is true. That hardly seems to count in favor of or to justify (to any degree) my going to the party, let alone to be what such a reason consists in. Contrast this with the case in which the fact *that there will be dancing at the party tonight* raises the chances that I'll have a *desire* satisfied. So far as I can see, this last proposal seems no better situated to constitutively explain reasons than the first. Neither seems to capture what it is for a fact to recommend, favor, or justify my going to the party. Otherwise put, **Reason** does not bridge what I earlier called the *practical/explanatory divide* (at least when this divide pertains not just to concepts but also to properties).

The Structure of Normative Facts

My project in the preceding section was to raise some general concerns about attempts to reduce moral features to descriptive ones. Jackson's strategy, we saw, not only yields bizarre results but is also insensitive to determination relations that may hold between properties or facts. Schroeder's proposal, in contrast, gives us no particular reason to believe that specifically moral features can be reduced to descriptive ones, and also appears not to satisfy the criteria for a successful reduction, being too revisionary in character. To point to the liabilities of these reductive proposals is, of course, not to establish that no reductive proposal could succeed. But the failure of these views, I believe, provides grounds for pessimism regarding future proposals.

In this section, I want to move from critiquing attempts by naturalists to reduce the moral to the natural to presenting a positive case for non-naturalism. As I indicated at the outset of our discussion, this case hinges on observations about the structure of normative facts: what moral facts would have to be like if they were to exist.

My point of entry into this topic is a recent discussion by William FitzPatrick (2008), in which FitzPatrick develops his own version of non-naturalist realism. He observes that naturalists such as Jackson have often developed their views by attempting to identify moral features with massive disjunctions of natural properties. I have pressed the point that such views are insensitive to determination relations that might hold between properties or facts. FitzPatrick takes the argument a step further, specifying the types of determination relation that hold when something is a moral fact.

Like Schroeder, FitzPatrick begins with a thought experiment. Consider, says FitzPatrick, an ordinary artifact such as the computer on your desk. Now ask yourself why that computer is a good one. Your answer will appeal to ordinary natural properties, such as its processing speed, its propensity not to crash, and so forth. For ease of reference, call these features of your computer XYZ. Though your computer is good in virtue of having these properties, its being good does not consist solely in its having them, for any number of things could have these properties without being good. Moreover, when it comes to artifacts, at least, good-making characteristics are relative to kinds; sharpness is a good-making quality in a knife, for example, but not in a coffee cup. FitzPatrick maintains that this point is important

> not because…[it] translates directly to the ethical case – it does not, as at least some ethically good- or bad-making properties may plausibly have that status invariantly…but because it directs us to the more complex structure of facts about goodness, which does have an analog in

> ethics. In the present case, the point is that the fact that this computer is good consists not only in the fact that it has XYZ, but in this *together with* the fact that XYZ is such as to satisfy the *standards of goodness* S for computers. (2008, 186)

By noting that the goodness of an artifact of some kind is determined by the standards of goodness for members of that kind, FitzPatrick does not wish to defend non-naturalism about such standards. Rather, he wishes to direct our attention to, as he puts it, the structure of goodness-involving facts, or as I would prefer to put things, the structure of normative facts.

FitzPatrick's argument appeals to parallel cases: To account for the normative properties of artifacts, we must appeal to normative standards, such as the standards of goodness for computers or knives, which explain how their natural properties contribute to their goodness or badness *qua* artifact. Likewise, to account for the moral properties of actions, we need to appeal to moral standards, which explain how an action's natural properties contribute to its having such properties as being morally permissible or impermissible.

Let's develop FitzPatrick's point. Imagine being present at a social gathering in which a senior colleague, who has a strong sadistic streak, publically humiliates a junior colleague with a clever but vicious put-down. Your senior colleague's behavior is morally wrong. But it is wrong not simply because it causes humiliation and social tension, which are ordinary natural states or properties, but also because there is the moral standard that:

It is wrong to humiliate others simply for pleasure.

I would add that this last point about the structure of moral facts extends to other cases, such as those involving etiquette and prudence. To see this, imagine that the same senior colleague who has acted badly has also acted rudely and irrationally. If his action is rude, however, then it is in virtue of his having violated standards of politeness that apply to his behavior. And if his action is practically irrational – say, because it deeply alienates those whose company he generally enjoys – then it is in virtue of his having violated applicable standards of prudence. This suggests that the structure of normative facts to which FitzPatrick draws our attention is pervasive; it is found not simply in the moral domain but also in the normative domains quite generally (cf. 2008, 188).

In sum, FitzPatrick defends three claims:

(1) There are particulars, such as act-tokens and individual people, that have moral properties. A particular's having a moral property is a particular moral fact.
(2) When a particular has a moral property, it has that property in virtue of there being some moral standard which that particular satisfies or fails to satisfy.
(3) Moral standards are general moral facts.

While the first claim is accepted by nearly all moral realists, the other two are more contentious. For that reason, I'll revisit them in a moment. Before I do, however, it's worth asking this: Why would a view according to which there are moral standards that determine particular moral facts favor non-naturalism in any regard?

Perhaps the best way to see why is to work with a pair of contrasting cases. Consider a game such as ice hockey. Like most games, hockey is a normed activity, since it's governed by rules or norms for how to play it. For example, the rules or standards that govern the game tell us that a standard game is divided into three periods, at most only six players on

each team can play at once, a player is not allowed to enter the opponents' zone before the puck does, the puck is considered "dead" during play if it leaves the perimeter of the rink, and so on. Along with these standards are rights, responsibilities, and obligations that attach to participants in the game, namely, players, coaches, and officials. Players, for example, have a right to take a shot on goal but they do not have the right to determine whether such a shot counts as a goal (officials do); officials have the right to determine whether a player is off-side but coaches do not; officials have the responsibility of tracking how many penalties a player might have but players do not; and so on. These rights, responsibilities, and obligations are enforced in various ways. If a player, for example, fails to comply with them, he or she may be penalized, assessed a fee, or suspended. If this is right, the standards that govern the game of hockey play at least two roles: they determine not only whether someone is playing the game correctly or incorrectly but also whether an activity counts as hockey at all. For were some activity that had the trappings of hockey – a game played on a rink with sticks, a puck, and so forth – not to sufficiently conform to the standards of the game, this activity would not count as a game of hockey.

To my knowledge, no philosopher has advocated non-naturalism regarding the standards that govern hockey and the correlative rights, responsibilities, and obligations that attach to participants in the game. Still, it will be worthwhile to think about the ontological status of the standards themselves, if only to understand better why these standards have been assumed to be naturalistically acceptable. We can do this by asking whether the standards of hockey bear various "marks of the natural."

According to many, the paradigmatic mark of the natural is to be fitted to play an explanatory role in the sciences. This is why quarks, DNA, and nitrous oxide are rightly deemed natural: they bear this mark. The standards of hockey, however, certainly don't bear this mark, since no science would advert to these standards to explain anything in the natural world. Another mark of the natural to which some philosophers appeal when defending the naturalistic credentials of some entity is being causally efficacious. Once again, though, the standards of hockey don't seem to bear this mark. In fact, if such standards exist, they look, for all the world, to be not particular concreta that are causally efficacious, but abstract entities that exist but are not spatially located (though they might have a history if we brought them into existence). If these standards enter into causal explanations at all, they will do so only by being the contents of our propositional attitudes.

So far, the standards of hockey don't seem to bear the marks of the natural. But notice this: these standards have the property of being deeply contingent. By this I mean that not only is the existence of the game of hockey contingent, since we can easily imagine never having invented it, but also the game itself could (at least to some degree) be governed by different standards. We could imagine, for example, the National Hockey League modifying the game in such a way that during certain periods of the game, seven players from a team are allowed to play at once.

There is an explanation for the deep contingency of the standards of hockey: they owe their existence or implementation to our activity. It is we who implemented these standards, by performing acts of authorization. For, at some point, authorized parties performed speech acts in which they pronounced that hockey games would be composed of three periods, that no more than six players from one team would be allowed to play at the same time, and so on. When they pronounced these things, they implemented the standards that govern the game. (I don't mean to suggest that this is the only way to implement standards. It is probably one among several.) These standards, moreover, continue to be in force.

That is because we continue to care about them and enforce penalties when they are violated. If we were to cease to care about or enforce the standards of hockey, these standards would probably still exist, but they would fail to govern anyone's activities.

In making these observations, we are still a considerable distance from adequately explaining how the standards of hockey are or could be constitutively explained by paradigmatic natural features. In fact, I believe that there are important unresolved puzzles regarding how these standards could be so explained (see Cuneo 2014). My purpose in discussing the standards of games such as hockey, however, is not to build a case for holding that we should be non-naturalists about them; rather, it is to compare them to moral standards, which non-naturalists believe we do have reason to hold are not constitutively explained by paradigmatic natural features.

To that end, let us assume that the concerns raised earlier about views such as Jackson and Schroder's are weighty enough to place these broadly reductionist positions to the side. If we do, then we need to ask whether moral standards are plausibly viewed as being descriptive, bearing the marks of the natural. We've already noted that, according to many advocates of naturalism, the paradigmatic mark of the natural is being such as to play an explanatory role in the natural sciences. Are moral standards fitted to play such a role? On the face of things, no; moral standards seem no better situated to play such a role than the standards of hockey, since our best theories in biology, physics, chemistry, and the like do not advert to properties such as *being wrong* or *having a moral right*.

Naturalists such as Nicholas Sturgeon suggest, however, that even if moral facts are not such as to play explanatory roles in the natural sciences, they are nonetheless plausibly viewed as being causally efficacious (Sturgeon 1988). For example, if we wished to address the question why your senior colleague acted badly, we might advert to moral facts such as his being morally callous. In my view, Sturgeon might be correct to say that some moral facts, such as *that your colleague is morally callous*, could causally explain happenings in the world (see Cuneo 2006; see also Shafer-Landau 2003; Wedgwood 2007). Still, it would not follow from this that the moral standards themselves causally explain anything. Like the standards of hockey, moral standards look for all the world not like concreta, such as quarks or concatenations of chemicals, but like abstracta, which are neither spatially nor temporally located. As such, if they enter into the causal nexus, it is because we represent them in certain ways.[5]

It is at this point, however, that we come to a crucial difference between the standards of hockey and the standards of morality. Unlike the standards of hockey, many of the standards of morality do not look contingent at all. Instead, they look to hold necessarily, at least for beings like us in a world such as ours. If we hold fixed our present constitution and environment, it looks as if the following standards *must* hold:[6] It is wrong to humiliate others simply for pleasure; is wrong to kill someone simply because she has inconvenienced you, and; is wrong to engage in recreational slaughter of fellow persons. One explanation of why they must hold is that it belongs to the essence of wrongness that killing someone simply because she has inconvenienced you is wrong. If such a view were correct, it would have important implications for how we could know moral standards, for we could know these standards by grasping important features of the essence of properties such as *being wrong*. If this were so, however, then our knowledge of the standards would not be via

5 A full vindication of this claim would, however, have to engage with Railton (1986), who claims that normative standards can enter into nomic explanations even when not the content of our beliefs.

6 This is compatible with these facts having *ceteris paribus* clauses built into them. More on this in a moment.

wholly empirical means or through the usual methods of the sciences, such as drawing generalizations by inductive means from particular cases. It would involve something like *a priori* insight into the way things are.

In sum, according to the argument presented in this section, there are particular moral facts, such as the fact *that your senior colleague's behavior is wrong*. These particular facts exist, however, in virtue of there being moral standards, such as the standard that it is wrong to humiliate others simply for pleasure. A close look reveals that standards such as this do not bear the marks of the natural. If things are as they appear, then that is good reason to believe that these standards are not natural.

The argument I've rehearsed in this section is subject to several objections. Let me close by voicing three reservations about it.

Earlier I noted that realists such as FitzPatrick hold that:

> Moral standards are general moral facts.

While most realists would accept this claim, there are some who would deny it. For example, David Copp holds that moral standards are not facts but imperatives, which are neither true nor false; nonetheless, these imperatives ground particular moral facts. If Copp is right, the moral standard concerning humiliation mentioned earlier is really an imperative whose content is explicitly expressed by the sentence, "Don't humiliate others simply for pleasure!" (Copp 2007, 13–14).

While I do not wish to deny that imperatives can be used to express moral standards, I see no reason to believe that such standards are imperatives if realism is correct. Familiar considerations adduced against traditional forms of expressivism (which hold that the contents of moral judgments express pro or con attitudes and are not true or false in any robust sense) tell against such a view. It seems to make perfect sense, after all, to say such things as, "I know that it is wrong to humiliate others simply for fun," and, "It is a fact that it's wrong to humiliate others simply for fun," and so forth. Moreover, we seem to be able to embed propositions that have standards as their content in logical schemata such as modus ponens. If Copp's view were correct, however, then these sayings and embeddings would not make perfect sense. To make sense of them would require considerable work – work that realists have thought both unnecessary and ultimately futile (see Schroeder 2008). While not decisive, these considerations, I believe, weigh heavily against the view that standards are not facts but mere imperatives.

The second reservation I shall consider raises a different concern. When developing the argument in this section, I appealed to the claim that:

> When a particular has a moral property, it has that property in virtue of there being some moral standard which that particular satisfies or fails to satisfy.

Some realists, however, reject this claim because they hold that that there are no moral standards. Somewhat more specifically, these philosophers advocate a view known as *particularism*, which assumes that moral standards would have to be truths, and that there are no true and useful moral standards.[7] At best, according to particularists, there are moral

7 Those familiar with the literature on particularism know that its advocates, such as Dancy (2004), present it in different ways. The description I offer of the view, I believe, captures its essence.

"rules of thumb." These rules of thumb are as some think of Newton's laws of physics: strictly speaking false, but helpful for getting around the world.

Why would particularlists hold that there are no true and useful moral standards? The primary argument that these philosophers voice is one that appeals to the holism of reasons. Roughly stated, the idea is that a descriptive feature such as *being such as to humiliate simply for pleasure* is not *itself* a wrong-making property. Whether it is depends on contextual features. In some contexts, this descriptive feature provides powerful moral reasons not to humiliate; in other contexts, such as when we're playing a game that involves humiliating other contestants, it does not. If this is so, we cannot claim that there are unqualified moral standards that are at once true and useful.

This last claim strikes me as correct. However, as best I can tell, absent other assumptions, little of interest follows from it. What nonparticularists should say is that most moral standards exist and are implicitly qualified by *ceteris paribus* clauses; these standards are, in ordinary conditions, true for the most part. Admittedly, specifying those conditions in which, say, humiliating merely for pleasure is not a wrong-making feature might not be easy, since those conditions would be fairly remote from ordinary experience. Nonetheless, this is compatible with these standards being true. And it would not follow that, if moral standards were qualified by *ceteris paribus*, they would be unhelpful when engaging in moral deliberations. To the contrary, they might be indispensable.

I turn to a final concern. Toward the beginning of our discussion, we distinguished between concepts, on the one hand, and properties and facts, on the other. Some (and perhaps all) propositions, I suggested, are composed of concepts, while some facts are composed of properties. Naturalists, I observed, might claim that while moral propositions are composed of non-natural moral concepts, these concepts characterize only natural properties or facts. Throughout our discussion in this last section, however, I have (mostly) spoken as if moral standards are facts. But one might deny this, claiming that such standards are propositions composed of non-natural moral concepts, while the properties to which they refer are natural. On the further assumption that such truths would hold in virtue of the essences of their constituent non-natural concepts (and not simply because of some further fact that makes them true), this view would imply that while there are non-natural moral truths (since they are composed of and made true by non-natural moral concepts), there are no non-natural moral facts. Instead, all facts are natural.[8] This view, then, would deny FitzPatrick's claim that:

Moral standards are moral facts.

In my view, this last objection is instructive for several reasons. For one thing, if it were correct, it would imply that the divide between certain versions of naturalism and non-naturalism might not be as wide as many have believed. After all, if the position we're considering is correct, there is a thesis that both naturalists and non-naturalists can accept: namely, that there are non-natural moral truths.

Perhaps more importantly, however, the objection we're considering illustrates that the argument offered for moral non-naturalism in this section is incomplete. It is true that we have uncovered no natural facts that are plausible candidates for being moral facts. Still, non-naturalists have more work to do if they are going to satisfactorily defend the claim

8 This is the position that Cuneo and Shafer-Landau (2014) call *minimal moral non-naturalism*.

that moral properties are non-natural and that moral standards are non-natural facts (as opposed to true propositions constituted by non-natural moral concepts). My own view is that non-nonaturalists can accomplish this work in a variety of ways, including by arguing that only facts can play the role of being that which explains why particulars have moral properties, as (2) asserts. However that may be, a benefit of our discussion is that we can better see exactly where more progress needs to be made.

Acknowledgments

Thanks to Kelly Clark and Russ Shafer-Landau for their comments on this chapter.

References

Copp, D. (2007). *Morality in a Natural World*. Oxford: Oxford University Press.

Cuneo, T. (2006). Moral Facts and Configuring Causes. *Pacific Philosophical Quarterly* 87: 141–162.

Cuneo, T. (2007). Recent Faces of Moral Nonnaturalism. *Philosophy Compass* 2: 1–32.

Cuneo, T. (2014). *Speech and Morality*. Oxford: Oxford University Press.

Cuneo, T. and Shafer-Landau, R. (2014). The Moral Fixed Points: New Directions for Moral Nonnaturalism. *Philosophical Studies* 171(3): 399–443.

Dancy, J. (2004). *Ethics Without Principles*. Oxford: Oxford University Press.

Enoch, D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.

FitzPatrick, W. (2008). Robust Ethical Realism, Non-Naturalism, and Normativity. In *Oxford Studies in Metaethics*, vol. 3, edited by Russ Shafer-Landau, pp. 159–206. Oxford: Oxford University Press.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Gibbard, A. (2012). *Meaning and Normativity*. Oxford: Oxford University Press.

Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Jackson, F. (2012). On Ethical Naturalism and the Philosophy of Language. In *Ethical Naturalism: Current Debates*, edited by Susana Nuccetelli and Gary Seay, pp. 71–88. Cambridge: Cambridge University Press.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

McPherson, T. (2012). Mark Schroeder's Hypotheticalism: Agent Neutrality, Moral Epistemology, and Methodology. *Philosophical Studies* 157: 445–453.

McPherson, T. (2015). What is at Stake in Debates among Normative Realists? *Noûs* 49(1): 123–146.

Olson, J. (Forthcoming). The Metaphysics of Reasons. In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star. Oxford: Oxford University Press.

Parfit, D. (2011). *On What Matters*, vol. II. Oxford: Oxford University Press.

Plantinga, A. (2010). Naturalism, Theism, Obligation and Supervenience. *Faith and Philosophy* 27: 247–272.

Railton, P. (1986). Moral Realism. *Philosophical Review XCV*: 163–207.

Scanlon, T.M. (2014). *Being Realistic about Reasons.* Oxford: Oxford University Press.

Schroeder, M. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.

Schroeder, M. (2008). *Being For*. Oxford: Oxford University Press.

Shafer-Landau, R. (2003). *Moral Realism*. Oxford: Oxford University Press.

Sturgeon, N. (1988). Moral Explanations. In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, pp. 229–255. Ithaca: Cornell University Press.

Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.

29

Naturalism and Moral Psychology

CHRISTIAN B. MILLER

This chapter considers a certain strand in recent work in ethics that falls under the broad heading of naturalism and moral psychology. I will not mean anything very precise by "naturalism" here, other than to be clear that I only have in mind methodological versions of naturalism, rather than ontological ones. Such methodological versions typically start by claiming that the sciences provide us with the best approach to understanding reality. In contrast, ontological versions of naturalism make claims about what does and does not exist, the latter including such entities as the theistic God, angels, and ghosts.

The strand running through naturalistic discussions is one that takes seriously empirical work which pertains to moral psychology, and from that starting point ends up drawing certain conclusions in metaethics. In particular, the conclusions are ones that go against traditional moral realist positions, or positions which claim that moral facts and properties exist objectively.[1]

In this chapter, I consider the work of four leading naturalistic moral psychologists: Joshua Greene, Shaun Nichols, Jesse Prinz, and John Doris. Each of them draws a different metaethical conclusion, and they would likely disagree among themselves on a number of points. But here my goal is to consider, as much as space allows, whether the moral realist should feel threatened by the empirical work that they cite and the arguments that they base upon it.

From Dual-Process Theory to Noncognitivism

In one of the most controversial research projects in contemporary moral psychology, the psychologist Joshua Greene has outlined an argumentative strategy that takes us from a dual-process theory of moral judgments to conclusions in metaethics which favor specifically noncognitivist forms of antirealism.[2] To focus the discussion, I will concentrate on

1 For a detailed discussion of how to formulate moral realism, see Miller (2009a).

2 This section draws on Miller (2009b).

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

Greene's 2008 paper "The Secret Joke of Kant's Soul," which expands on the metaethical implications of his work in psychology to a greater degree than his other articles.

The central move Greene makes is to argue that what he calls "deontological" moral judgments tend to be caused by emotional responses and are *post hoc* rationalized, whereas so-called "consequentialist" moral judgments tend to be caused by more cognitive processes and involve genuine moral reasoning (2008a, 36). Consequentialist and deontological judgments are defined functionally in terms of characteristic moral conclusions, such as "Better to save more lives" in the first case and "It's wrong despite the benefits" in the second (2008a, 39). On Greene's view, each of these ways of thinking stems from distinct psychological patterns in the same person's brain, patterns that have a long evolutionary history (2008a, 37). One pattern is cognitive, and is connected to the dorsolateral surfaces of the prefrontal cortex and parietal lobes. The other is emotional, and is connected to other parts of the brain, like the amygdala and the medial surfaces of the frontal and parietal lobes (2008a, 40–41).

Schematically, it is possible to hold that consequentialist and deontological judgments are both primarily cognitive, or that they are both primarily emotional, or again that consequentialist judgments are more emotional and deontological ones more cognitive. Greene opts for the remaining fourth option, and in particular advances the hypothesis that "what deontological moral philosophy really is, what it is *essentially*, is an attempt to produce rational justifications for emotionally driven moral judgments, and not an attempt to reach moral conclusions on the basis of moral reasoning" (2008a, 39, emphasis his). He summarizes a range of different empirical studies to support his hypothesis, which I will only briefly mention here:

- *Trolley Cases:* To explain the differences in typical reactions to the standard trolley case and the footbridge variant, Greene claims that the thought of having to push someone to his death is more emotionally salient than, for instance, merely having to flip a switch in the standard case. Increased neural activity is found in emotional-response regions of the brain in footbridge cases, whereas in trolley cases increased activity is found in the more cognitive regions (2008a, 43).
- *Singer Cases:* Similar results are found in cases involving a nearby drowning child versus cases involving impersonal donations (2008a, 47).
- *Victim Cases:* Similar results are found where the contrast is between identifiable victims such as Baby Jessica, who is trapped in a well, and indeterminate, merely statistical victims (2008a, 48).
- *Punishment Cases:* Deontological, retributivist judgments about punishment tend to be emotionally driven and indeed are "proportional to the extent that transgressions make [people] angry" (2008a, 51). But when discussing punishment in the abstract, people often give consequentialist arguments.
- *Harmless Action Cases:* Cases involving harmless actions, such as certain instances of breaking a promise, are often condemned as a result of emotional responses, whereas in more reflective moments, people are less willing to forbid them (2008a, 55).

Deontological moral judgments are thus taken to be causal offshoots of our moral emotions, whereas consequentialist judgments are more inherently cognitive (2008a, 63–64).

The second half of Greene's paper concerns the philosophical implications he draws from these results. Unfortunately, the discussion is condensed, and several arguments get

run together here. Thankfully, Mark Timmons, in a commentary article, has nicely distinguished and formulated four different arguments (2008, 95–102). Here, I want to highlight the two that bear most directly on moral realism:

- *The Coincidence Argument:* Given that deontological moral judgments seem to be emotional gut reactions for which there is a good evolutionary explanation, the "rationalist" deontologist must explain how there could be such a coincidence between these emotional reactions and the rationalist's posited objective moral truth. As Greene says, "it is unlikely that inclinations that evolved as evolutionary by-products correspond to some independent, rationally discoverable moral truth" (2008a, 72).
- *The Sentimentalist Argument:* Deontology is committed to the view that moral judgments are (or are expressions of) cognitive states like beliefs, but in light of the empirical data, a noncognitivist, sentimentalist story is the way to go.

Thus, according to Greene, these arguments should lead us away from moral realism and toward a noncognitivist position (2008a, 76).

Mark Timmons not only helpfully clarifies Greene's arguments, but also suggests forceful responses.[3] Let us look in a little more detail at the exchange between Timmons and Greene concerning the coincidence argument. Timmons notes (rightly) that there is no entailment relation between adopting a deontological approach to normative theory and being a realist about moral truth. For instance, Scanlon's constructivism serves as a nice example of a nonrealist foundation to a deontological ethic (Timmons 2008, 98). And on constructivist accounts, it would not be much of a coincidence if the moral truth were to reflect to some degree our emotional intuitions, since those very intuitions would be suitably refined by the requisite process of rational reflection.

Greene has replied to Timmons' commentary, and specifically on this point he thinks that deontologists who adopt a constructivist approach would fall prey to a dilemma. Either the emotionally based deontological intuitions that go into the reflective process would also come out of it, in which case we have the result of "garbage in, garbage out." Or they would not come out, in which case the resulting moral claims are not necessarily deontological (Greene 2008b, 116).

Even though I am no constructivist, it seems clear that much more needs to be said here. Take the first horn. If the methodology and principles used in carrying out the reflective process are themselves rationally supported (and perhaps epistemically acquired via reflective uses of our "cognitive" brain centers), then other things beings equal, and when properly utilized, they should bestow justificatory authority on any deontological intuition that they legitimize. To take simple examples, the Kantian categorical imperative procedure associated with the first formulation, the Rawlsian veil of ignorance, and the Scanlonian reasonable rejection criterion could each allow us to systematically refine our initial deontological intuitions so that the end product is no longer "garbage out," but rather specific moral commitments that now have substantive justificatory backing. And this leads to the second horn of the dilemma. While Greene is certainly correct in thinking that the outputs of the constructivist procedure might not be the exact same initial intuitions and so might not *necessarily* be deontological in form themselves, he gives us no reason to think that they *would not* be either. Indeed, at least on certain interpretations of their projects, Kant, Rawls,

3 For an extremely careful and thorough evaluation of Greene's arguments, see Berker (2009).

and Scanlon have all applied their own constructivist procedures and generated from them recognizably deontological principles and moral evaluations of cases.

These remarks even open the door for a deontological *moral realist* to respond to Greene as well. For Greene offers us no reason, at least in this particular discussion, to distrust the workings of the more cognitive centers of the brain. But then, a moral realist might provide a story about how reason can be used to properly discern one or more general objective *deontological* moral principles, principles which in turn can be used as action guides and also as checks on our specific intuitions. Those intuitions that are cleared by the rationally discovered objective moral principles would then be given substantive justificatory backing. Note that nothing about the empirical results Greene provides has any bearing on the metaphysical existence of such objective moral principles or, as far as I can see, on the psychological possibility that our cognitive faculties can gain epistemic access to them.[4]

So, for now, this particular path from results in moral psychology to moral antirealism seems murky at best.

From Sentimental Rules to Moral Relativism

Perhaps a clearer path can be discerned if we start from the account of moral judgments offered by Shaun Nichols in his 2004 book *Sentimental Rules*.[5] Nichols is not out to provide a conceptual account of moral judgments, but rather an empirically informed account of what we are doing when we make what he calls "core moral judgments." Such judgments are concerned with the moral permissibility of actions, and in particular whether such actions have violated harm-based norms. Judging the wrongness of an act of hitting an innocent person purely for amusement would be a paradigm example of such a core moral judgment (2004, 5).

There are two central components to Nichols' account of core moral judgments (in what follows, "core" will be dropped for the sake of simplicity). The first is a person's normative theory, which is his or her body of information about what actions are wrong. And the second is an affective mechanism that is responsive to harms, where these are understood mainly as pain and suffering. The first component is used to explain why some events can be distressful and involve pain and suffering but are not considered wrong, such as natural disasters and accidents (2004, 15). The need for a second component is motivated by Nichols in various ways using cases involving psychopaths, autistic children, and disgust. For instance, psychopaths have a difficult time with the moral/conventional task: they do not make a distinction for moral norms on the basis of permissibility, seriousness, and authority independence, and give conventional rather than welfare-based justifications for both kinds of wrongdoing. This failure to form core moral judgments is linked by Nichols to another failure of psychopaths, namely their significantly lower physiological responses to distress cues (2004, 12–13). So the idea is that psychopaths often understand both harm and conventional norms, but have

4 A similar strategy exists for deontological moral realists who are skeptical about there being just one (or a few) foundational objective moral principles, such as certain moral pluralists and particularists.

Very helpful in this context is also Joyce (2008), who nicely distinguishes between three different kinds of moral rationalism and notes that two of them – conceptual and justificatory rationalism – are not likely going to be threatened by the kinds of neuroscientific result that moral psychologists such as Greene offer.

5 This section and the next draw on material from Miller (2011).

neither backed by an affective system, whereas normal subjects have an affective system linked to their harm norms. Together, then, for Nichols these two components of a normative theory and affective system give us an account of actual moral judgments as involving "sentimental rules," or "rules prohibiting actions that are independently likely to elicit strong negative affect. The set of rules or normative theory prohibits actions of a certain type, and actions of that type generate strong affective response" (2004, 18).

Even if we accept that there are these two components to moral judgments, what we need is a detailed account of how they work together and what precise relation they bear to core moral judgments. And here Nichols is short on details, remarking at one place that they "somehow conspire to produce the distinctive responses tapped by the moral/conventional task" (2004, 29). Nichols goes on to say a little more about their relationship, but not nearly enough in my view to make it very clear how this "conspiring" is supposed to happen. Similarly, he seems to reject the claim that both components operating together occurently are necessary for core moral judgments (2004, 28–99). Instead, it may be enough if the affective system is present at some crucial earlier developmental stage (2004, 29). But even so, we are never given an argument for why such an affective system is *necessary at any point* for the production of core moral judgments, where the necessity here is presumably nomological rather than conceptual necessity, since Nichols is not giving a conceptual account of moral judgments. Instead, at best, what his arguments for the role of an affective system would seem to show is just that such a system is extremely common or frequently present in subjects making core moral judgments. Furthermore, even if both components are indeed necessary in some way, Nichols says nothing to convince us that they are jointly *sufficient*. And so, without any clear reasons to accept either the necessity or the sufficiency of his view, it is not clear to what extent we have an actual empirical *account* of our capacity to form core moral judgments on the table yet. Finally, even if Nichols provides such an account, it is initially unclear how it would extend beyond just those judgments concerning violations of harm-based norms.[6]

For the remainder of this section, however, let me leave aside these worries and examine what bearing, if any, Nichols' account would have on metaethical discussions about the objectivity of morality. Nichols himself is clear that he takes it to provide evidential support for a relativist position. Indeed, the conclusion of his central argument is that "[n]o action is wrong simpliciter. At best, an action is only wrong relative to a population – the population of individuals that share a certain emotional repertoire" (2004, 185). Here is the argument itself:

(1) Rational creatures who lack certain emotions would not make the moral judgments that we do.
(2) There is no principled basis for maintaining that these certain emotions (on which our moral judgments depend) are the *right* emotions. That is, there is no externally privileged basis for maintaining that all rational creatures *should* have the emotions.
(3) [Therefore, morality is not objective.] (2004, 185, emphasis his)

The opponent here is the philosopher who claims that the moral status of an action is determined "as it is in itself" and so holds that moral judgments are true in a nonrelativistic manner (2004, 184). And while this might be the folk understanding of the status of morality,

6 For similar concerns, see Sinclair (2005).

Nichols takes the argument to show that the "commonsense commitment to objectivity is unwarranted. Given the emotional basis of moral judgment, we are not justified in our belief that morality is objective" (2004, 185).

It is hard to know what to make of this argument, since for one thing it is clearly not logically valid as stated, nor is it obvious how best to supplement it with additional premises. But let me instead focus on the premises that we do have. Nichols' main argument for the first premise is that "Martians who lacked analogues of human sentiment and affection would not make the moral judgments that we do" (2004, 185). Indeed, the Martians might be aware of all the same relevant facts, but judge that torturing puppies is not morally wrong (2004, 185). Given the sentimental rules account, affect systems play a crucial role in shaping both the initial moral judgments that we make and the ones that as a culture we preserve over time. Without those affect systems, the resulting judgments would look very different.

There seem to be several problems with this supporting argument for premise (1). First of all, it relies on the necessity of affective systems for the capacity to form core moral judgments, but as we noted earlier, Nichols has not provided an actual argument for their necessity as opposed to their mere commonality. But without such an argument, his first premise should not be stated definitively in terms of what such rational creatures without affective systems would or "would not" judge. Second, even if affective systems are necessary, recall that Nichols is only providing an *empirical* account of the capacity for core moral judgments, and presumably one that applies just to *human* moral judgment capacities, rather than to moral judgment capacities as such. So, while it might be true that non-human rational creatures who lacked certain emotions might not make moral judgments *in the way that we do*, it is a much further step, and one that seems hard to defend given only what has been said so far, to claim that such creatures would not arrive at moral judgments with the same content as ours *in some other way*. Finally, whether Martians would or would not make such a judgment about torturing puppies is presumably an empirical matter. And this is not meant to be a trivial point: premise (1) is stated strongly in terms of what certain non-human rational creatures *would not* do, but how can this be determined ahead of time? Rather, at best, all that Nichols is entitled to here is the claim that such creatures *might not* or *possibly* would not make the same judgments that we do. But then it is unclear what happens to the rest of the argument once such a weaker premise is used.

Admittedly, if the first premise is simply trying to capture the idea that many moral judgments tend to vary based upon underlying affective sensitivities, then it could be readily accepted. As Nichols notes, emotional deficiencies in psychopaths can explain why they do not form core moral judgments, and when it comes to nonmoral disgust norms, variations in disgust-sensitivity can lead to variations in the authority contingency of norm violations, and perhaps even to variations in the patterns of cultural evolution for etiquette norms (2004, 186–187).

So let us proceed to premise (2), which seems to be the core of Nichols' argument. Nichols explains this premise more fully as follows: "there is no principled basis for maintaining that all rational creatures should have emotional responses like reactive distress and concern. There is no independent reason to think that this emotional repertoire is the right one to have" (2004, 187). Instead, the burden of proof is supposed to be on the moral realist to "show that all rational creatures *should* have such emotional responses…The difficulty is that it is not at all clear how to argue for such a claim" (2004, 188, emphasis his). Note, then, that Nichols does not seem to provide any actual support for premise (2), other than just

taking it to be the default position and putting the burden of proof on the realist to justify certain emotions over others.

At this point, a realist about moral facts might raise the following concerns about premise (2), and more generally about Nichols' argument:

(i) Premise (2) seems to preach to the choir. A philosopher who is already committed to some version of moral realism presumably has good (albeit perhaps mistaken) reasons for countenancing the existence of objective moral facts, and presumably would be willing to claim that a proper subset of those facts concerns the appropriateness of forming certain emotional responses rather than others. So Nichols has offered nothing by way of an argument for why the *realist* should accept this premise, and rather just seems to be asserting that it is true.

(ii) Related to this first point is a second one. The claim in premise (2) that "there is no principled basis for maintaining that these certain emotions are the right emotions" appears to be, not a psychological claim, but rather a metaphysical one about the existence of objective normative standards. But recall that Nichols is offering an empirical account of the psychology of core moral judgments. And by itself, nothing follows from such a view about the metaphysics of objective moral facts – indeed, by itself it is completely neutral on such metaphysical issues. For even if our affective systems completely determined the content of our subjective normative theories (which is highly unlikely), that would still have no bearing on the metaphysical existence of such facts. It might change how we think about the epistemic status of our normative theories, but it should not change how we think about the metaphysical issues themselves. Thus, the motivation for premise (2) understood as an ontological claim about objective standards has to come from somewhere other than the sentimental rules account itself, and as far as I can see that motivation is not to be found in Nichols' book.

(iii) Finally, it is not clear more generally why a principled basis for justifying certain emotions over others cannot be found in the agent's own normative theory. Recall that on the sentimental rules account, the capacity for core moral judgments is constituted by an agent's normative theory and affective system. And, presumably, our normative theories could provide us with a wealth of important reasons for why, say, emotions sensitive to distress in others are worth having whereas emotions promoting distress in other are not. One such reason, for instance, could be that relieving distress in others is important (for various reasons of its own), and beings with an emotional sensitivity to distress in others are more reliable at detecting that distress and so are better equipped to try and relieve it. And Nichols would not be entitled to respond at this point by claiming that such reasons provided by our normative theories are themselves arbitrary and have "no principled basis," since that would assume there are no objective reasons to begin with and so beg the very question at issue against the moral realist. Nichols might reply that this response underestimates the impact that affective systems have on normative theories. That is, he might hold that affective systems strongly influence the contents of the normative theories we hold in the first place, especially at the cultural level, where they significantly impact what moral claims survive over time. Indeed, Nichols devotes an entire chapter of his book to a cultural evolutionary account of harm norms, in which affective systems play a central role (2004, ch. 7).

Here I could imagine the realist making use of one of two strategies to respond. The first might be to argue that the best explanation for cultural changes in harm

norms over time, and by implication the best explanation for the harm norms the majority of us have today as part of our normative theories, does not involve appealing primarily to our affective systems, but rather to human beings having made progress in getting closer to the objective truth about morality. To his credit, Nichols anticipates this response, and argues against it at length (2004, 159–164). At this point, I leave it to the reader to see who gets the better of that exchange.

The other strategy worth briefly mentioning is to concede that human affective systems have played a significant role in shaping our normative theories, but to note that they are only one of several factors which causally *influence* those theories, rather than being the only factor that causally *determines* them. For Nichols has not offered arguments that come anywhere close to showing that the content of our normative theories is determined *solely* by our affective systems. Furthermore, we could take seriously the distinction between causal and justificatory influence, and note that while the affective system only contributes the former in shaping an agent's normative theory, other forces might provide both kinds of influence. For instance, through a process of practical reasoning, an agent might come to a new moral conclusion about the need to send money to Africa or the inconsistency between her moral views on two issues. Such processes are familiar, and yet often contribute increased justificatory status to their conclusions. Or, as another example, the influencing factor could be the social transmission of accumulated wisdom about moral topics through forms of testimony such as parental upbringing and education, some of which corresponds to the objective moral facts. And, indeed, there might be cases of such social transmission leading to a change in normative theory even in the face of resistance from the individual's affect system. To use one of Nichols' own examples, "[k]nowing that inoculations are for the best does not eliminate the discomfort one feels witnessing a child get inoculated" (2004, 155; see also Vranas 2006, 789).

Stepping back, the upshot of this discussion of Nichols' argument is that the sentimental rules account of our capacity for core moral judgments might not force us in the direction of moral relativism after all, and indeed seems to be compatible in a number of respects with a realist metaphysic about moral facts and properties. Thus, while Nichols initially appeared to offer us a clear path from a naturalistic moral psychology to moral relativism, such an appearance is deceiving.

From Constructive Sentimentalism to Moral Relativism

In his 2007 book *The Emotional Construction of Morals*, Jesse Prinz develops a constructive sentimentalist view which also tries to provide us with a clear path to moral relativism. As we will see, it is a form of moral relativism that is particularly hard to accept.

Prinz begins by advocating what he calls "emotionism," or the view that emotions are essential to morality. This approach comes in a metaphysical form as follows:

- *Metaphysical Emotionism:* Moral properties are essentially related to emotions (2007, 14).

Such a view accepts the existence of moral facts and properties, and so is not a nihilist metaethical position. At the same time, by making morality dependent on the emotions,

it falls short of the kind of mind-independence typically accepted by moral realists. Rather, on this view, morality exists, but in such a way as to be constructed by the emotions (2007, 14).

To have epistemic access to moral facts, one needs corresponding moral concepts. And if those facts are constituted by emotions, then the moral concepts are likely bound up with emotions as well. Hence, we get another version of emotionism:

- *Epistemic Emotionism:* Moral concepts are essentially related to emotions (2007, 16).

But one can hold one form of emotionism without the other. For instance, classical utilitarians are metaphysical but not epistemic emotionists. On the other hand, traditional noncognitivists reject metaphysical but accept epistemic emotionism (2007, 17).

A strong emotionist theory accepts both metaphysical and epistemic emotionism, and perhaps the leading example of such a view is moral sentimentalism (2007, 20). This is the direction in which Prinz himself chooses to go. Schematically, sentimentalist views take the following rough form:

- *Metaphysical Thesis:* An action has the property of being morally right (wrong) just in case it causes feelings of approbation (disapprobation) in normal observers under certain conditions.
- *Epistemic Thesis:* The disposition to feel the emotions mentioned in the metaphysical thesis is a possession condition on the normal concept RIGHT (WRONG) (2007, 20–21).

Such an approach countenances the existence of moral properties, and defines them as powers to cause emotions in us (2007, 89). They exist as real causal features of the world, but as relational features, much as colors are often understood as secondary qualities of physical things. Situations have these powers to cause relevant emotions in us, and such moral properties can ground the truth or falsity of moral statements and can be quantified over in explanations of behavior. At the same time, moral facts are not natural scientific facts, but rather social facts, like monetary values, constructed by our psychological dispositions (2007, 167). Furthermore, at the psychological level, moral properties are represented by the emotions they cause in us.

The metaphysical and epistemic theses only characterize the broad outlines of a sentimentalist approach in general, rather than Prinz's own constructive sentimentalism. Here is how he initially refines the metaphysical thesis for moral wrongness:

(W) An action has the property of being morally wrong just in case it causes feelings in the spectrum of both self-blame and other-blame emotions in normal observers under certain conditions (2007, 90).

Such blame emotions include shame and guilt (self-blame) and anger, contempt, and disgust (other-blame). Clearly, (W) needs further development: "normal observers" and "certain conditions" are not very illuminating clauses. In order to bypass having to give an account of either, Prinz instead invokes his technical term of "sentiments," which are dispositions to have occurent emotions (2007, 84). This leads to the following settled proposal:

(W*) An action has the property of being morally wrong (right) just in case there is an observer who has a sentiment of disapprobation (approbation) toward it (2007, 92).

So talk of "normal" observers has dropped out of the picture, and the "certain" conditions are just those in which an observer has the sentiment in question toward the action.

From here, we can provide a corresponding refinement to the epistemic thesis. On Prinz's view, we said that moral concepts are constituted by emotions, and so the moral concept WRONG gets characterized as follows:

(E) The standard concept WRONG is a detector for the property of wrongness that comprises a sentiment which disposes its possessor to experience emotions in the disapprobation range (2007, 94).

More straightforwardly, when an agent has the belief that some action is wrong, this amounts to having "a sentiment of disapprobation toward it" (2007, 94). In other words, the agent has a long-term memory representation (the disposition), which disposes her to feel either guilt or shame if she is the one performing the action (self-blame) or either anger, contempt, or disgust if someone else is (other-blame). Such a sentiment is, for Prinz, a moral rule, and the manifestation of that sentiment in a particular case is a moral judgment (2007, 96).

At this point, an example would help, and Prinz provides a nice one (2007, 96). Suppose you observe a pickpocket taking a wallet. A series of mental events may naturally result. First the pickpocket's action is interpreted as a case of stealing. Additionally, you are likely going to have a sentiment in your long-term memory toward stealing, which constitutes a rule. When you observe the pickpocket's action and classify it as theft, it may cause the sentiment to be activated and a particular emotion to be formed, depending on the contextual factors involved. Guilt and shame are not appropriate, since another person is performing the action – rather, anger seems like a natural candidate in many people. So, you are angry at the pickpocketing, and since emotions constitute moral concepts and moral judgments, this anger in part constitutes the moral judgment that pickpocketing is wrong (2007, 96).

How does Prinz's constructive sentimentalist view differ from Nichols' sentimental rule view? The central difference concerns the relationship between moral judgments and emotions. As we have just seen, for Prinz, emotions constitute moral judgments as component parts in virtue of constituting moral concepts. On the other hand, we know that, for Nichols, normative judgments and affect systems are distinct and can exist independently of each other in a number of human beings (such as psychopaths). Prinz offers several reasons for preferring his own approach over Nichols', but we do not need to dwell on them for our purposes here (2007, 99–100).

There are many other complexities to Prinz's view. For instance, he devotes an entire chapter to providing a subtle noncognitive theory of the emotions, and then uses it to better understand moral emotions. But, in the remainder of this section, let me focus specifically on the connection Prinz draws between constructive sentimentalism and moral relativism. And this connection is meant to be robust: Prinz claims that moral relativism is a "straightforward consequence" of his view (2007, 173). Why? Here is the argument:

(1) Descriptive moral relativism is true.
(2) If descriptive moral relativism and constructive sentimentalism are true, then metaethical moral relativism is true.
(3) Therefore, metaethical moral relativism is true (2007, 174–175).

If constructive sentimentalism is true, then the rightness or wrongness of an action depends on an individual's moral sentiments. And if descriptive moral relativism is true, then different people have different views on fundamental moral issues. So, together, the two views imply that "the existence of differences in people's sentiments entails a difference in moral facts" (2007, 175). Prinz has already supported constructive sentimentalism at length, and by drawing on anthropological data concerning such practices as Roman gladiatorial games and Inuit infanticide, he provides a sustained defense of premise (1) as well (2007, 187–195). Hence the "straightforward consequence" of metaethical moral relativism.

Prinz accepts a version of individual as opposed to cultural moral relativism, and holds that terms such as "wrong" and "ought" have an implicit indexical element which determines the content of an utterance contextually in part as a function of the values of the speaker making it (2007, 181). Thus, an utterance such as "Slavery is wrong" is, on Prinz's account, shorthand for "Slavery is wrong according to the values of the appraiser or some other salient individual(s)" (2007, 200). In addition to articulating and motivating his version of metaethical moral relativism, Prinz also goes to great lengths to defend it from objections.

In the remainder of this section, I want to briefly raise two concerns about Prinz's moral relativism. The first is that it seems to deliver too few moral properties in certain cases. And the second is that it seems to deliver far too many moral properties in other cases.

To motivate the first concern, consider a possible world with human beings organized in a *Brave New World* society consisting only of ruling elites together with enormous numbers of workers. The ruling elites have sentiments of approbation toward the social system they have devised and the specific abusive ways in which they treat their workers as slaves, while the workers have been genetically programmed and culturally conditioned from birth to love the elites and fully endorse their own position in life. Recall that on Prinz's view:

(W*) An action has the property of being morally wrong just in case there is an observer who has a sentiment of disapprobation toward it.

But, in this case, neither the elites nor the workers have such sentiments. Hence, it seems to follow that none of the abusive actions of the elites toward the workers would be wrong. Even worse, (W*) would imply that they would be *right* (and indeed, given what Prinz says about obligation (2007, 179–181), even *obligatory*), both from the perspective of the elites and from that of the workers. But surely, many of us would say, such practices are still wrong in that world despite what people there might happen to think?

Prinz seems to suggest two different ways of treating cases like this. The first is to rigidify our moral terms such as "wrong" so that we can still say the institution of slavery is wrong in such a world, where the use of "wrong" here is grounded in *our* current sentiments (2007, 149). Admittedly, this would allow the statement that slavery is wrong in that world to come out true, but I suspect this response is not likely to satisfy many critics. For, while it is an empirical question worth further study, I imagine that many people believe that slavery is wrong in such a world *independently* of whether we are in a position to judge that it is wrong or not. Rather, from the perspective of that very world (rather than from the perspective of our sentiments in this world), the statement that slavery is wrong is true.[7]

Elsewhere, Prinz seems instead to simply bite the bullet. When considering a world of psychopaths who regularly cheat and kill, he writes: "What are we to say about murder in

7 For more on the rigidifying strategy, see Miller (2009a).

this community? Is it morally wrong? I don't think so. It's no more immoral for a psychopath to kill, in this horrible world, than it is for a lioness to kill a wildebeest on the African savanna" (2007, 130). And one proposed explanation for this is that since psychopaths do not have moral sentiments, they do not fall under the purview of moral rules (2007, 130).

Whatever may be the case for a world of psychopaths, I doubt many of us would want to adapt this response to the slavery world. First of all, humans in that world very well could have moral sentiments such as anger, disgust, and guilt. It is not that they do not *have* moral sentiments, but rather that such sentiments have been radically *corrupted*. Second, by analogy, we would have to say that the hierarchical society in that world would be no more immoral than, say, the hierarchy found in an ant society. And that is certainly a tough pill to swallow.

So Prinz's account seems to deliver too few moral properties in some cases. But it also seems to deliver too many in others. For recall that on his settled proposal (W*) about the metaphysics of moral properties, an action is right or wrong depending on the sentiments of "an" observer. And consider an action such as ordering millions of people to be sent to the Gulag. One observer was Stalin, and another was one of the innocent people sentenced to hard labor. Given (W*), such an action would seem to have the property of being both right and wrong, presumably relative to different frameworks (though this qualification is not part of (W*), which is formulated in terms of wrongness *simpliciter*). And we can also throw in good, bad, virtuous, vicious, blameworthy, commendable, and the whole range of moral properties – the action will instantiate all of them (excluding, perhaps, the deontic properties of obligatory, merely permissible, and wrong, for the reasons Prinz gives in 2007, 175–180). While the observer has to be real, as opposed to imaginary, and has to actually have the sentiments (understood as real mental dispositions), as opposed to merely having a passing feeling, such restrictions have no bearing on cases such as this.

This consequence of (W*) is a specific instances of a larger concern which Prinz does raise, namely that the "problem is not that moral claims are false, but rather that too many moral claims are true" because of the "thousands of moralities throughout the world" (2007, 288). And given all of these moralities, an action such as Stalin's will be assessed in different ways, with no approach seemingly available to the relativist for justifying one assessment over another, besides the parochial strategy of just using his or her own relative sentiments. Nor can any robust claim to moral progress be made when, for instance, the Soviet work camps are closed (2007, 288–289).

To deal with these concerns, Prinz provides a long list of normative criteria to use in evaluating competing moral systems, such as consistency, reliance on false factual information, promotion of social stability, ease of implementation, generality, and impact on welfare (2007, 289–292). Such criteria are intended to be nonmoral, and so do not help us decide whether a given moral system is more moral or closer to the truth than any other system (2007, 292). At the same time, they allow for a kind of progress: we can make nonmoral progress when we shift to a moral framework that comes out ahead, other things being equal, on one of these criteria.

Let me end this section by very briefly raising some concerns with this proposal:[8]

(i) One might think that Prinz's proposal fails to capture what we care about with the notion of progress. It seems that when people say we have made progress in eliminating slavery or ending genocide in our country, part of what they mean is that we have

8 For similar concerns, see Joyce (2009, 517–518).

made *moral* progress in such a way that we are a better nation morally as a result. But Prinz has only offered an account of nonmoral progress.

(ii) More troublesome is the fact that the criteria Prinz has proposed are all normative criteria, and presumably lend themselves to a sentimentalist story as well (it would be strange to pick and choose one's sentimentalism about norms!). But then, combined with a descriptive relativism about such criteria, metanormative relativism looms here, too. So *exactly* the same kinds of worries about the proliferation of normative properties, the difficulties of comparative assessment, and the coherence of normative progress will arise in this area as well. Assessing one set of relative norms with another set of equally relative norms hardly seems to advance the discussion.

(iii) Finally, it is not clear that many heinous moral frameworks and practices (by our lights) would not come out just fine on Prinz's criteria. Genocidal practices such as Stalin's could be argued by their practitioners to be highly coherent with other values of order, equality, economic advancement, and the like, while also in their opinion being easy to implement, leading to greater social stability, applying generally to a range of cases, and so forth.

Prinz is aware of concerns such as these, but ultimately resorts to the fallback response of claiming that his main goal was just to "explain how progress is possible *at all*" (2007, 300, emphasis his). But that seems like a weak response. For surely it is easy from a moral relativist perspective to show that it is *possible* to adjudicate between competing moral systems and develop an account of progress simply by putting forward some normative standard (whether moral or not) that is true only relative to your own system of values, and then go from there. Maybe that is all we can reasonably expect moral relativists to be able to offer.

So here, too, the path from a naturalistic moral psychology to antirealism seems fraught with pitfalls.

From Social Psychology to an Error Theory about Character

Let me end this chapter by highlighting an alleged connection between naturalistic moral psychology and, not noncognitivism or moral relativism, but rather an error theory about our moral judgments.[9] In contrast to the previous sections, here I am much more sympathetic to the work that is being done, in part because it has only modest metaethical consequences.

Over the past 15 years, one of the central topics in ethics has been the implications of social psychology for the empirical adequacy of character traits. The leading contributors have been Gilbert Harman, in a series of articles, and especially John Doris, in his 2002 book, *Lack of Character*. In this section, I will focus on Doris's line of argument, as it is more thoroughly developed.

Doris's target is what he calls a *globalist* conception of character, which is one that accepts the following two theses:

- *Consistency:* Character traits are reliably manifested in trait-relevant behavior across a diversity of trait-relevant eliciting conditions that may vary widely in their conduciveness to the manifestation of the trait in question.

9 This section draws on material from Miller (2014, ch. 7, 8).

- *Stability:* Character traits are reliably manifested in trait-relevant behaviors over iterated trials of similar trait-relevant eliciting conditions Doris (2002, 22).[10]

A *global character trait*, then, is a character trait that exhibits both cross-situational consistency in a wide variety of trait-relevant circumstances and stability in repeated instances of the same kind of trait-relevant circumstances. To take an example, someone who is courageous is expected to exhibit courage in a wide variety of relevant situations (i.e., the battlefield, the courtroom, the sports field, etc.), as well as in repeated instances of the same kind of situation (i.e., many battles over multiple years).

With the target clarified, Doris argues as follows:

(1) If there is widespread possession of the traditional virtues and vices understood as global character traits, then systematic empirical observation using appropriate psychology experiments will reveal most people behaving in a certain kind of way.[11]
(2) However, systematic empirical observation using appropriate psychology experiments fails to reveal that most people act in this kind of way.
(3) Therefore, there is not widespread possession of the traditional virtues and vices understood as global character traits.[12]

This "behaving in a certain kind of way" is acting virtuously in a wide variety of situations relevant to the particular virtue in question. Thus, the main focus of Doris's argument is the cross-situational consistency of traditional virtues and vices, and he claims that there is little empirical evidence that most people are consistently virtuous from one relevant situation to the next.

The specific virtue that ends up being central to both Harman and Doris's discussion is compassion. They reason that if most people were compassionate, then at the very least they would perform a variety of simple helping tasks, and we could reliably predict when they would probably help in the future. But, in study after study, a significant number of participants do not help, and our predictions about their behavior are often badly mistaken.[13]

The studies that Harman and Doris appeal to are familiar ones in psychology, such as the Darley and Batson Princeton Theological Seminary hurry study, the Milgram shock

10 Doris also mentions a third globalist thesis, evaluative integration (1998, 506; 2002, 22), but I leave it to one side in what follows, as he does in much of his discussion. In a recent article, talk of "global character traits" is replaced with talk of "robust character traits" (Merritt, Doris, and Harman 2010, 356). The terminology of "global" traits is not original to Doris; it has been used for decades in psychology.

11 Doris (1998, 523, fn. 23).

12 While he does not formulate the argument this explicitly in either Doris (1998) or (2002), something like this seems to be what he has in mind. See in particular Doris (1998, 505–507) and Merritt, Doris, and Harman (2010, 357–358).

13 This does not quite capture their concern, though. For, in some studies, a majority of participants in one group *do* indeed help. But then there will be another group in which very few people help. The only relevant difference between the two groups is a morally insignificant change in the situation, such as a change in the smell or temperature. These are not changes that should make a difference to compassionate people. But they clearly result in significant differences in behavior. Hence, most people do not have the virtue of compassion, because as premise (1) indicates, they are not behaving in the *kind of way* that people with such a virtue would.

Thus, "both disappointing omissions and appalling actions are *readily* induced through seemingly minor situations. What makes these findings so striking is just how *insubstantial* the situational influences that produce troubling moral failures seem to be" (Merritt, Doris, and Harman 2010, 357, emphasis theirs). See also Doris (1998, 507; 2002, 2, 28, 35–36) and Harman (2003, 90).

experiments, the Isen and Levin dime phone booth study, and the Latané and Darley group effect studies. All of these studies are intended by Harman and Doris to bear on the extent to which people have the specific virtue of compassion. But they assume, albeit without citing much supporting evidence, that similar empirical results can be found which challenge the widespread possession of other traditional virtues, as well as the vices.

It is worth getting clearer about what exactly the Harman/Doris conclusion is supposed to be. It is not anything like:

(i) On metaphysical grounds, the properties of *being compassionate* and *being honest* do not exist.

That would not be the kind of claim that psychology experiments could establish. Furthermore, they are not arguing that:

(ii) No human being has ever had any of the traditional virtues or vices, such as courage or compassion, either as a matter of psychological necessity or as a matter of contingent fact.

Indeed, Doris in several places concedes that the psychological evidence is compatible with a few people having the virtues and a few other people having the vices (Doris 2002, 60, 65, 112, 122; see also Vranas 2005, 16).

On the other hand, they clearly are not just after as weak of a claim as:

(iii) Given the psychological evidence, we are not justified in believing on the basis of that evidence that most people possess the traditional virtues or vices.

While they would readily accept this, it is only a claim about the absence of evidence, namely that we currently lack support from psychology for the widespread possession of these traits. But Harman and Doris repeatedly make stronger claims than this based on their reading of the psychology literature. Indeed, the title of one of Harman's early papers is "The Nonexistence of Character Traits" (2000), and he claims that "it may even be the case that there is no such thing as character" (1999, 328). Doris argues that "people typically lack character" (1998, 506; 2002, 2).[14] And together they have written that "Behavior is *not* typically ordered by robust traits" (Merritt, Doris, and Harman 2010, 358, emphasis theirs).

So instead, I think the right interpretation of their conclusion is this:

(iv) Given the psychological evidence, we are justified in believing on the basis of that evidence that most people *do not* possess the traditional virtues or vices.

Let me make two observations about this conclusion.

First, note how modest it is from a metaethical perspective. Since it does not bear on the metaphysical existence of moral properties concerning character, it is entirely compatible with a robust moral realist position. In other words, the property of honesty could exist

14 Similarly, speaking for the situationist position that he supports, Doris writes that the situationist "denies that people typically have highly general personality traits that effect behavior manifesting a high degree of cross-situational consistency" (2002, 39; see also 6, 64).

objectively, and just be rarely instantiated. So in this sense, no antirealist conclusions follow from Harman and Doris's project.

Second, I do not think we should challenge (iv). Indeed, in two recent books, I have argued in much greater detail that (iv) is a conclusion we have *excellent* reason to accept.[15]

Harman and Doris use (iv) as a stepping stone to a second line of argument directed at the plausibility of Aristotelian virtue ethical accounts, along with that of any other theories in normative ethics which rely centrally on global traits. This part of their project will not be my concern here.[16] Instead, let me end by spelling out a bit of the connection to error theories in metaethics.

In general, an error theory typically starts with a claim about something a class of people believes:

(B) Certain people have beliefs whose content involves some feature, F.

For instance, moral error theories typically take the form of a cognitivist claim that:

(B1) Most ordinary people form moral beliefs whose content involves some moral feature, F.

This claim is opposed to traditional noncognitivist positions, which instead understand moral judgments as expressions of desires. Note also that the first claim of error theories is psychological – it pertains to what the people in question are *thinking*. The second claim, on the other hand, is *metaphysical*:

(M) F fails to obtain or to be instantiated.

So the moral error theorist will add the metaphysical claim that:

(M1) Moral feature F fails to obtain or to be instantiated.

Hence, the conclusion of combining the psychological claim with the metaphysical claim is that the people in question have erroneous thoughts in this area of their lives:

(C) Certain people have beliefs whose content involves F, and those beliefs are false.

Most error theories in metaethics have focused on the axiological (e.g., goodness, badness) and deontological (e.g., obligatory, optional, forbidden) concepts, with little attention paid to character concepts such as the moral virtues and vices (see Mackie 1977; Joyce 2001; 2006). But there is a natural way to extend Harman and Doris's work into a version of an error theory, as follows.

Let me start with this initial claim, using the example of the virtue of compassion:

(B*) On the basis of observing helping behavior, most people infer (whether consciously or not) that certain individuals – perhaps family members, friends, community leaders, politicians, or the like – are compassionate people.

15 See Miller (2013; 2014).

16 See Miller (2014, ch. 8) for extensive discussion.

To this we can add the metaphysical claim that:

(M*) Few people actually have the virtue of compassion.

Thus, it follows that:

(C*) Most of our actual moral judgments involving the ascription of compassion to one or more individuals turn out to be false.

Less formally, think about the people in your life whom you think are compassionate. Perhaps they include some of your friends, or a leader in the community. If the argument for (C*) is correct, then it is likely that these people do not really have the virtue of compassion. So, while you might think you know their character well, many of your beliefs in this area actually turn out to be mistaken.

Nor would the error theory just stop with compassion. The argument can be generalized so that it applies to *any* traditional moral virtue, such as courage and honesty, as well as to all the traditional moral vices, such as cowardice. So the conclusion really is that:

(C**) Most of our actual moral judgments involving the ascription of either traditional moral virtues (compassion, honesty, nonmalevolence, etc.) or moral vices (cold-heartedness, dishonesty, cruelty) to one or more individuals turn out to be false.[17]

Call this particular version of a moral error theory the *character error theory*.

The first premise of the argument, generalized to all the traditional virtues and vices, is a familiar observation from ordinary life. It seems that most people attribute virtue and vice concepts not just to actions, and not just to motives, but also to people themselves. We tend to think that our friends are honest, or loyal, or trustworthy, for instance. Indeed, Bernadette Park examined weekly descriptions by a group of strangers who got to know each other during a 7-week period, and found the "most striking result from these data was the tremendous prevalence of traits, accounting for 65% of all the information, followed by behaviors (23%)" (1986, 910). Park also found that trait information offered by the participants increased in prevalence over time, as they got to know one another better.[18]

The second premise is just the conclusion that Harman and Doris have advanced, and that I have also supported in my own work. Indeed, my view is that the experimental evidence from psychology supports the claim that most people have character traits which are neither virtuous nor vicious (see Miller 2013; 2014).

Hence, I think we should take seriously a robust error theory about our ordinary moral character trait ascriptions, but one which, as already noted, does nothing to threaten a moral realist outlook about character properties.[19] Harman and Doris were wise not to claim otherwise, as this potential path to moral antirealism is a dead end.

17 For statements of something like this claim, see Harman (1999, 316, 329), Doris (1998, 513–514; 2002, 1071–1108), and Vranas (2005, 29).

18 Furthermore, Ostrom found that, when asked what information would be helpful in forming an impression of another person, participants gave responses involving character traits more often than responses involving beliefs or behaviors (as reported in Park 1986, 908).

19 For more on the character error theory, see Miller (2014, ch. 7).

Conclusion

This chapter has briefly reviewed and evaluated four different and influential research projects that, in recent years, have tried to draw philosophical implications from a naturalistic moral psychology. Other approaches are also worth considering in a longer discussion.[20] My only claim is that as of now, the views we have considered so far do not seem to give us much reason to favor an antirealist approach to the status of morality.

Acknowledgments

Thanks to Kelly James Clark for inviting me to contribute to this volume. Some of the material draws from Miller (2014) and is reprinted with permission from Oxford University Press. Additional material draws from Miller (2011). Support for this work was funded in part by a grant from the Templeton World Charity Foundation. The opinions expressed in this paper are my own and do not necessarily reflect the views of the Templeton World Charity Foundation.

References

Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy and Public Affairs* 37: 293–329.

Churchland, P. (2012). *Braintrust: What Neuroscience Tells Us about Morality*. Princeton: Princeton University Press.

Doris, J. (1998). Persons, Situations, and Virtue Ethics. *Noûs* 32: 504–530.

Doris, J. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Greene, J. (2008a). The Secret Joke of Kant's Soul. In *Moral Psychology,* vol. III: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, pp. 35–80. Cambridge, MA: MIT Press.

Greene, J. (2008b). Reply to Mikhaul and Timmons. In *Moral Psychology,* vol. III: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, pp. 105–118. Cambridge, MA: MIT Press.

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108: 814–834.

Harman, G. (1999). Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society* 99: 315–331.

Harman, G. (2000). The Nonexistence of Character Traits. *Proceedings of the Aristotelian Society* 100: 223–226.

Harman, G. (2003). No Character or Personality. *Business Ethics Quarterly* 13: 87–94.

Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.

Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

Joyce, R. (2008). What Neuroscience Can (and Cannot) Contribute to Metaethics. In *Moral Psychology,* vol. III: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, pp. 371–394. Cambridge, MA: MIT Press.

20 Such as the work of Jonathan Haidt (2001) or Patricia Churchland (2012). I discuss Haidt's work in Miller (forthcoming).

Joyce, R. (2009). Review of Prinz, J. *The Emotional Construction of Morals. Mind* 118: 508–518.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.

Merritt, M., Doris, J., and Harman, G. (2010). Character. In *The Moral Psychology Handbook*, edited by J. Doris and the Moral Psychology Research Group, pp. 355–401. Oxford: Oxford University Press.

Miller, C. (2009a). The Conditions of Moral Realism. *Journal of Philosophical Research* 34: 123–155.

Miller, C. (2009b). Review of Walter Sinnott-Armstrong, *Moral Psychology*, vol. 3: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Notre Dame Philosophical Reviews. Available from: http://ndpr.nd.edu/news/24102/?id=16786 (last accessed July 15, 2015).

Miller, C. (2011). Moral Relativism and Moral Psychology. In *The Blackwell Companion to Relativism*, edited by Steven Hales, pp. 346–367. Oxford: Blackwell.

Miller, C. (2013). *Moral Character: An Empirical Theory*. Oxford: Oxford University Press.

Miller, C. (2014). *Character and Moral Psychology*. Oxford: Oxford University Press.

Miller, C. (forthcoming). Social Intuitionism versus Morphological Rationalism: Assessing Two Competing Approaches to the Psychology of Moral Judgments. Philosophical Explorations.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Park, B. (1986). A Method for Studying the Development of Impressions of Real People. *Journal of Personality and Social Psychology* 51: 907–917.

Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Sinclair, N. (2005). Review of Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Notre Dame Philosophical Reviews. Available from: http://ndpr.nd.edu/review.cfm?id=4261 (last accessed July 15, 2015).

Timmons, M. (2008). Toward a Sentimentalist Deontology. In *Moral Psychology,* vol. III: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, pp. 93–104. Cambridge, MA: MIT Press.

Vranas, P. (2005). The Indeterminacy Paradox: Character Evaluations and Human Psychology. *Noûs* 39: 1–42.

Vranas, P. (2006). Review of Nichols, S., *Sentimental Rules: On the Natural Foundations of Moral Judgment. Mind* 115: 784–790.

30

Militant Modern Atheism

PHILIP KITCHER

1

In times when violence carried out in the name of religion abounds, when many groups of people seek to interfere with the private lives of others because those targeted are allegedly violating divine commands, and when important discoveries about the world in which we live are questioned, or even denied, because they are supposed to be incompatible with authentic messages from the deity, it is easy to think that things have gone too far. Polite respect for odd superstitions about mysterious beings and their incomprehensible workings might be appropriate so long as the misguided folk who subscribe to them do not seek to convert, coerce, or eliminate outsiders, but, when the benighted believers invade the public sphere, it is important that they not be earnest. Further, respect should not extend to the deformations the faithful exert upon the minds of the young: just as children deserve to be protected against parents who refuse to allow them to receive medical attention, so too are they entitled to defense against forms of religious education that will infect and corrupt their abilities to think clearly and coherently. We no longer inhabit the arcadias of Waugh and Wodehouse, in which fanatic believers and their aggressive challengers who ask where Cain found a wife are equally figures of fun. Because of religious belief, our world is an oppressive and dangerous place, and it is time for those who value reason, justice, tolerance, and compassion to do something about it.

Militant modern atheism, whose manifesto I have just summarized, is an eminently comprehensible reaction to features of public life, perhaps most evident in the United States and in parts of the Muslim world. Nobody ought to deny that the writings of some of those who articulate the principal themes in this manifesto, particularly Richard Dawkins (2006) and Daniel Dennett (2007), contain cogent arguments and genuine insights. Though many of the arguments and conclusions are anticipated in the works of earlier writers – in d'Holbach, Hume, and Russell, for example – it is important to have

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

them restated, eloquently restated, in the context of contemporary human conditions and in the light of current knowledge.[1]

Yet the militant response to the dangers and oppressions that are so easily traceable to religious zeal has elicited a further reaction. The books of the "four horsemen" (of whom Dawkins and Dennett are two), who see themselves as riding to the defense of a world besieged by threatening nonsense, have received severe, sometimes savage, critiques from reviewers who regard them as adopting a pinched and distorted conception of religion and of its role in human life.[2] These critics do not want to align themselves with those religious movements that provoke militant modern atheism. They too regret that crude and intolerant forms of religion exist. Their charge is that the self-styled horsemen are meeting primitive fanaticism with equally primitive fanaticism, so that distortions of a valuable human institution – religion – are mistaken for the healthy original. Militant modern atheism, so viewed, is itself akin to the fundamentalism it opposes: intellectually simplistic, aggressively intolerant, and dangerously polarizing. By reducing the options and posing a stark choice, people with reasonable religious views are likely to be led to perceive their own practices as threatened, and thus to make common cause with the zealots.

Militant modern atheists have replies to these charges. They can point out, reasonably, that the most prominent, and probably the most prevalent, forms of contemporary religion are not the subtle ideas cherished by their critics, but the cruder doctrines they directly attack. They can challenge those who believe in "reasonable religion" to specify more clearly just what commitments such types of religion entail: to declare in public what has been abandoned, and to stick to the declaration. They can demand that those who profess a more enlightened religion no longer provide cover for fanatics who take a simplistic view of the scriptures they share with the sophisticated.

So far, I have aimed to outline an important contemporary debate. As a secular humanist, who shares many of the conclusions of the militant modern atheists but also agrees with the critics that valuable options are being foreclosed, I shall try, in what follows, to work through the issues involved. It is useful to start slowly.

2

Many contemporary discussions of religion, especially those that concentrate on the three Abrahamic monotheisms, distinguish fundamentalist forms from others that are more sophisticated. Fundamentalists are typically conceived as people who insist on reading the sacred texts as literal truth. Their more subtle co-religionists are distinguished by a willingness to declare that significant parts of the scriptures must be read as metaphors or allegories, or in the extreme case to suppose that the whole canon be interpreted nonliterally. Debates then center on whether one can defend the literal truth of *any important sentence* in the pertinent texts, on whether commitment to what can be maintained as literal truth can count any longer as religion, and on whether, if it can, that type of religion is at all

1 Dennett (2007) and Dawkins (2006) often recapitulate points made by Russell (1967). Some of the ideas and arguments go back to Hume's *Dialogues on Natural Religion* and Baron d'Holbach's voluminous antireligious writings (for example, *Christianity Unveiled*).

2 Thus, Leon Wieseltier has written and spoken scathingly about the books of both Dennett and Dawkins. Karen Armstrong (2009) also criticizes Dawkins' allegedly shallow understanding of religion.

representative of contemporary religious movements. Militant modern atheists will probably concede that some minor bits of the Bible can be counted as literally true – "Jesus wept," for example – but they repeatedly point out that allegorical and metaphorical readings are alien to the practices of the world's faithful. Like it or not, various grades of fundamentalism are omnipresent, and these are the proper targets of critique.

I want to reject this way of framing the debate for two main reasons. First, the notion of fundamentalism it introduces is not obviously coherent. Perhaps there are some sentences in the famous scriptural texts whose meaning is sufficiently clear and definite to allow one to speak of literal reading: to suppose, say, that "Jesus wept" is true in virtue of the fact that a particular historical figure shed tears. Plenty of the Bible and of the Qur'an is very different. What would it mean to interpret Paul's description of our condition literally, when he says that "we see through a glass darkly"? (The popularity of sunglasses in the ancient Near East?) What is the literal meaning of the references in *Genesis* to "days" at a time when the stars and planets are still being formed? Exactly what forms of struggle and resistance does the Qur'an commend to the faithful?

Historians of religion have pointed out the range of strategies of interpretation that have been popular at different periods, and have argued cogently that the idea of scripture as a scientific text, intended to meet the standards of clarity and precision toward which scientists strive, is – unsurprisingly – something that begins to emerge only with the unfolding of modern science, from the 17th century on. Like other readers, fundamentalists are up to their necks in the interpretation business, dedicated *in principle* to an enterprise of responding to the scriptures as if these works were aimed at an ideal of definiteness that typically seems alien to them, pursuing *in practice* lines of interpretation that have been laid down by leading figures in their particular religious tradition and that usually involve tendentious social and political commitments.

The second, and deeper, reason for not framing the discussion in the usual way lies in the fact that it centers the debate on the truth of claims found in particular documents. At first sight, that might appear to be a completely benign assumption. After all, aren't religions distinguished from one another by their various creeds? Defining "religion" is notoriously difficult, but one attractive strategy is to say that a religion consists of a set of doctrines about special kinds of entities ("transcendent" entities, different in kind from the everyday constituents of nature), that individual religions are distinguished by their different doctrines, and that to be committed to a particular religion is to believe the doctrines constitutive of that religion. Call this the *belief model* of religion.

As it stands, the belief model might seem inadequate, in that more than belief is required of the religious believer. Besides beliefs, there are emotions, aspirations, desires, and actions: devout Christians love God and their neighbors, aspire to accord with the divine commandments, wish that the Holy Spirit would descend on them and on others, work for the relief of suffering and the spread of the Gospel. Those who *merely* believe, if there are any such people, are not full participants in the religious life. There is a simple way to amend the belief model to encompass this judgment, namely to declare that belief is fundamental. The other psychological states of the religious believer flow from her beliefs. Believing the doctrines she does, including the claim that particular texts are true and represent the divine will, she is moved in particular ways, recognizes particular rules for conduct, forms her plans and goals, and, to the extent that her will is strong, expresses what she values in her actions. A friendly amendment to the belief model recognizes that religious believers have distinctive psychological states of several different types (emotions,

intentions, etc.), but takes belief in doctrine to be the fundamental basis from which the other religious states emerge.

This is *a* plausible way to think about religion and religious life, at least in the modern world, and perhaps especially in those parts of the world where movements that call themselves "fundamentalist" hold sway. It is, however, by no means the *only* conception of religion we might form. An alternative approach to religion can begin with other elements of the complex of states and processes, psychological and sociological, found in religious life, selecting some of those states and processes as basic, and viewing the distinctive doctrines as outgrowths, introduced as means of reaching goals marked out by the fundamental attitudes.

That formulation is both vague and general. My aim is to introduce a particular version that will exemplify the possibilities toward which I have just gestured, and I begin with the schematic suggestion to emphasize the fact that my specific conception is by no means the unique alternative to the belief model.

I begin with the concept of an *orientation*. A person's orientation identifies particular goals as valuable, not only with respect to his own life, but beyond the compass of that life. It contains a strong desire to work toward those goals, and, typically, to do so in coordination with others who share them. These desires support commitments to forming plans to promote the goals and to implement the plans in action, even if that should require considerable sacrifices and personal losses. An orientation, then, is a complex of psychological states – states of valuing, desires, intentions, emotions, and commitments – a complex that does not include factual beliefs, and that embodies a person's sense of what is most significant and worthwhile in his own life and in the lives of others.

Religious people whose psychological lives fit the belief model have something similar to an orientation, but in their case the complex of states that directs their conduct arises from specific beliefs about a transcendent entity. They typically suppose that traditional texts contain doctrines that express the will of this being, and, consequently, that particular goals are set for them. I am interested in considering forms of religious life for which the orientation is primary. People who participate in such forms of life have independent views about what is valuable, views they use to appraise the commands of any putative deity; they concur with Kant's claim that some independent source of value is required if "the Holy One of Israel" is to be assessed as good (1785/1981, 21). Some people for whom an orientation is basic will profess beliefs using sentences that others employ to make claims about the existence and character of transcendent beings (call these *doctrinal sentences*), and they may even share some beliefs with fundamentalists. For the religious people in whom I am interested, however, such beliefs are adopted because of the basic orientation: they are consequences of the commitment to particular goals and values.

Thoroughly secular people can have an orientation in the sense I have introduced. In their case, having their particular orientation gives rise to beliefs in neither transcendent beings nor any professions that use doctrinal sentences. Further, just as religious people can reflect on and thoroughly endorse the orientations they have, so too can their secular cousins. Let's say that an orientation is *reflectively stable* when it survives detailed scrutiny of one's life and the lives of others; when it can be upheld as a worthy choice for the direction of one's transient existence. The *orientation model* fits a human life when the person in question has an orientation and when that orientation is reflectively stable. This model can be applied to secular people and to people who consider themselves as religious. The latter are distinguished by the fact that their orientations lead them to particular forms of

behavior: to participate in particular ceremonies, to utter doctrinal sentences, and sometimes (perhaps often) to form beliefs whose content they express with such statements. Yet, when these people talk of their faith, they do not intend to classify their beliefs and professions of belief as psychological states that fall short of knowledge because of lack of evidence, but rather to single out the primary commitments made in their orientations. Faith is primarily a matter of thorough devotion to particular goals and values.

Within the sphere to which the orientation model is applicable, I want to distinguish four types, three of which are easy to appreciate. The *secular* are those for whom their orientation does not give rise to any beliefs about transcendent entities or any participation in professions and ceremonies associated with such beliefs. The *mythically self-conscious* are people whose orientation does lead them to utter doctrinal sentences and to participate in the professions and ceremonies in which those statements find their homes, but who clearly disavow any interpretation of the statements that implies substantive doctrine about transcendent entities. Thus, a mythically self-conscious Christian might describe herself as oriented by the values of human equality and solidarity expressed in particular Gospel passages (e.g., the Sermon on the Mount), so that she engages with other members of a Christian community to advance those values, while explicitly rejecting *any* interpretation of the New Testament that supposes a personal being from whose will those values derive, who in *any* sense created, governs, or surveys the universe. Her creedal professions are expressions of a commitment to the fundamental values and a supposition that, whatever the details of the history of the cosmos, these are the significant things for human beings to try to advance.

The *doctrinally entangled* go one step further. They hold some beliefs they recognize as implying the existence of transcendent beings, and they take these beliefs to be inspirationally important precisely because the pertinent beings exemplify the fundamental values. To the extent that the beings are viewed as especially powerful, they may even be seen as guaranteeing the possibility of realizing those values. A Christian who believes that Jesus was the incarnation of a being who created the cosmos can see himself as participating in a campaign to achieve important goals – the spread of loving relations among human beings – whose ultimate success is assured. The belief thus functions as a description of aspects of the universe and as an inspiring promise. If asked to defend his belief in particular claims about the transcendent, the doctrinally entangled person will not appeal primarily to evidence, but rather suggest that it is legitimate to form such beliefs because of the positive role they play in the promotion of the most important values.

Both the mythically self-conscious and the doctrinally entangled are clear about what they are doing when they utter religious statements. Between them lie characters whose ideas about how to interpret doctrinal sentences are far less definite. They are not prepared to say, with the mythically self-conscious, that there is *no* defensible interpretation of those sentences on which they are committed to the existence of transcendent entities. On the other hand, they are not willing to offer any definite interpretation that would provide a content to which they would subscribe. Many of them are inclined to take refuge in language that is resonant and opaque, metaphorical and poetic, and to deny that they can do any better at explaining the beliefs they profess. Something important happened, they think, at Mount Sinai, or in the desert outside Mecca, but they reject the idea that what occurred was any encounter between a historical figure – Moses, in the one instance, Muhammad, in the other – and a very large and impressive person with a long white beard or a somewhat smaller individual with wings and a gleaming presence. If pressed, they will admit that they can only gesture vaguely in the direction of something that might commit

them to the existence of transcendent entities – or might not. Their lack of definiteness frustrates militant modern atheists, who find no value in the resonant phrases that pervade theological discussions, but believers will contend that literal language gives out here, that as with great poetry, religious language somehow functions in ways that cannot be captured in the preferred modes of speech of their opponents. We can think of characters of this sort as the *doctrinally indefinite*.

At this point, I can state some theses I'll attempt to defend in the rest of this essay. First, militant modern atheism is entirely correct in its assault on those types of religious life that fit the belief model. On the other hand, all three of the nonsecular approaches that accord with the orientation model are defensible. In the case of the mythically self-conscious, that is hardly surprising, and the militant modern atheists applaud when those who continue to think of themselves as religious firmly reject "supernatural" entities – the militants think, however, that what remains hardly deserves the name of religion. More problematic, at first sight, are the cases of the doctrinally entangled and the doctrinally indefinite. I'll suggest that doctrinal indefiniteness can be a reasonable expression of epistemic modesty, and that even doctrinal entanglement can be justified when it is the only way of preserving, in the sociocultural environment available, a reflectively stable orientation. Militant modern atheism tends to overlook this point because it is in the firm grip of the belief model, and thus assumes – wrongly – that correction of belief about the occupants of the cosmos can automatically be articulated into a satisfying vision of what is valuable in one's life. Perhaps that is true for the privileged few, but it is not so for the less fortunate many.

Nevertheless, two important points emerge from the militant atheistic campaign. The first is that defensible religious attitudes cannot be tainted with elements of the belief model: that would occur, for example, if the beliefs justified by appeal to the promotion of values were then to serve as the bases for adding extra goals, viewed as expressions of the divine will. For the orientation model to apply, considerations of value must be primary, and the believer cannot contend, first, that commitment to the deity is legitimate because it expresses orientation toward things that are of fundamental worth, and, second, that the commitment brings in its train a host of new duties and prohibitions grounded in divine commands. The second point is that any liberal approach to the formation of orientations must recognize the importance of providing people with genuine choice – allowing them to find and pursue "their own good in their own way" (Mill 1859/2008, 15). This underscores the thesis, expressed forcefully by Dawkins (2006), that children should not automatically be assimilated to the religious practices of their parents.

3

To see why religious people who conform to the belief model are vulnerable to cogent critique, it is only necessary to note that they have two options for defending the legitimacy of their beliefs. The first is to claim that they have grounds for thinking that the doctrines they espouse are true, and in this case the dispute between them and their critics turns on matters of evidence. The second is to suppose that there are no such compelling grounds, that the ideas about transcendent entities are adopted on grounds that fall short – even very far short – of the evidence normally taken as required; in this case there are *ethical* concerns about how beliefs acquired by what some will call "leaps of faith" – and

others describe as "jumping to conclusions" – can serve as sources of values, goals, and actions in the ways the belief model supposes.

Though militant modern atheists sometimes delve into the "proofs" offered in premodern theology, these excursions are entirely irrelevant to the exploration of any possible evidence for doctrinal statements. Whether anyone has ever regarded them as a potential route from atheism or agnosticism to theism is doubtful, and it seems plain that those who constructed the more or less intricate arguments did so in attempts to elaborate ideas about a deity to whose existence they were already committed. People who fit the belief model come to the doctrines they espouse because of the social traditions in which they stand and because of the experiences (the "religious experiences") they undergo. Do these processes, individually or in combination, provide any warrant for doctrinal belief?

A sense of the presence of a deity (or of other sorts of being) is very common across the spectrum of the world's religions. Too common, in fact, to play any serious evidential role. Once you appreciate the widespread tendency of people to arrive at very different claims on the basis of experiences that seem to them both intense and mysterious, and see that these experiences are categorized in terms that derive from the religions with which the subjects are familiar – and to which they often subscribe – it is clear what is occurring. Religious and secular people alike experience things they cannot explain in everyday secular terms, and sometimes grope for categories that will make some sense of what has occurred. None of these assimilations is to be trusted, for none can maintain that it alone is privileged and that rival interpretations are erroneous. Nor will it do to apply some minimal category, to record one's encounter with some spectral person or some transcendent mind, for that, too, is one thought too many. Given the astonishing variety in the frequency with which religious experiences take place – the statistics suggest that there are "epidemics" of such experiences – and given also the correlations between religious experience and various other factors not conducive to cognition (ingestion of drugs, psychological distress, moments of adolescent ecstasy), trust in the deliverances of these experiences, conceived as routes to the transcendent that are quite independent of the support given by religious tradition, seems entirely unwarranted. Dewey offered wise counsel on these matters:

> when it [religious experience] occurs, from whatever cause and by whatever means, there is a religious outlook and function. As I have said before, the doctrinal or intellectual apparatus and the institutional accretions that grow up are, in a strict sense, adventitious to the intrinsic quality of such experiences. (1933, 17)

In other words, we should cherish the ways in which some experiences we do not fully understand reorient human lives, suggesting possibilities of promoting important values, without struggling to discern mysterious causes. To do that is to order one's life in the mode of the orientation model.

Religious people who exemplify the belief model are thus evidentially dependent on the traditions in which they stand. That is not yet to deny them the possibility of knowledge, since all of us are dependent on others for virtually all of what we know. The trouble is that the symmetry found in the appeal to religious experience – the use of phenomenologically similar episodes to support radically incompatible conclusions – is preserved when we turn to the grounding of belief in religious tradition. The Native American who is convinced of the existence of ancestral spirits and the Australian Aboriginal who talks confidently of the Dreamtime both base their religious professions on similar ideas about the past to those

that ground the doctrines of Jews, Christians, and Muslims. Once, long ago, there was a revelation, and it has been transmitted, with integrity, across the generations to the present. There is no reasonable way to break the symmetry, to declare that one – or some – of these supposed processes of revelation and accurate transmission have matters right, and the others are sad examples of primitive confusion. Moreover, when scholars study the processes through which the doctrines of major world religions evolve, and the ways in which those religions recruit converts, it becomes evident that they are shaped by causes we standardly view as unlikely to lead in the direction of truth. Nobody who recognizes the political considerations that have figured in the construction of the world's most celebrated religious texts can regard these scriptures as reliable indicators of past events. Nobody who reflects on what sociologists have to say about the ways in which people become attracted to particular religions will suppose that the spread of a creed has much to do with its truth.

Considerably more could be said about the difficulties of the belief model in sustaining the idea of religious knowledge, but I'll turn now to the alternative approach that conceives religious belief as unwarranted by the evidence. It is easy for opponents to formulate this approach as the supposition that faith intervenes to fill the gap between the inadequate evidence and confident belief, but that is to introduce a notion of faith quite different from the concept valued by many religious people. Devotees of one of the Abrahamic monotheisms whose lives are structured by the orientation model take their faith to be a form of commitment: they start from identifying themselves with particular goals and with a community that strives toward those goals, and that initial identification is perceived as the act of faith. Abraham can figure for them as the paradigm of this sort of commitment, as the "knight of faith," not because he leaps beyond the evidence to conclude that a specific transcendental being exists – that would be absurd, given his supposed personal encounters with this being – but because of his willingness to trust his God, even when perplexing and morally repugnant things are demanded of him. On the versions of the orientation account I shall defend, that sort of trust is not legitimate, for the commitments made are bound by ethical prescriptions (there ought to be no "teleological suspension of the ethical"), but the example serves as a useful correction to views of faith that remove it from the practical sphere and treat it as belief without sufficient evidence.

Any religious person who supposes that belief in transcendent entities can be legitimate without sufficient evidence faces the challenge of explaining how religious doctrines figure in the guidance of action. One possible response is to deny that they play any role whatever. They are placed in some sealed compartment in the believer's psychology, permanently unavailable for any practical decision-making. Whether or not this supposed insulation of religion is even coherent – that is, whether subjects whose minds were structured in this way could properly be said to believe the doctrines ascribed to them – I strongly doubt that any actual people satisfy these conditions. Yet, if the partitions between doctrinal belief and decision-making are permeable, then the presence of such belief is ethically suspect. Someone who makes decisions affecting the lives of others is ethically required to rely on those propositions best supported by the evidence. Thus, unless the doctrinal belief is better supported by the evidence than pertinent rivals, it will be wrong to use it in planning one's conduct. Since the evidential gap between belief about transcendent entities and the facts available to religious believers is typically a gulf, it is extremely hard to envisage how they could escape the charge of ethical irresponsibility.

Again, there are complications to be probed, but my principal interest lies in showing that *when the issue is framed in a particular way*, conceived in terms of the belief model,

militant modern atheists can make a powerful case. Frequently, they fail to present that case, thus inviting religious responses that rebut inferior forms of criticism. Yet, if the targets are forms of the religious life that accord with the belief model, skeptics can focus on the grounds of belief and pose a devastating dilemma.

4

Among the militant modern atheists, Dennett sees most clearly that the issues center on the processes that underlie religious belief. His efforts to articulate the theme are particularly instructive because they reveal how concentrating on the belief model can blind critics to possibilities that emerge from rival ways of thinking about religious life. Dennett provides a speculative evolutionary psychology, one that hypothesizes selective advantages in cognitive structures that inclined our ancestors to gravitate to certain types of explanation for puzzling phenomena, in particular to posit invisible agents as causes. The world's religions are unfortunate side products of a psychological tendency that served our forebears well. With the advance of modern science, people have better ways of responding to the puzzles, so that we can now liberate ourselves from faulty beliefs that have had all sorts of harmful practical consequences. We can "break the spell," self-consciously resisting modes of thought that, like our tendencies to folk physics, incline us to false beliefs.

Dawkins does not elaborate this theme with the detail and care that Dennett provides, but he shares a similar picture – hence the "God delusion." The common perspective is dominated by concepts drawn from the realm of knowledge and belief, framed by the belief model. Within this frame, the conclusions militant modern atheists draw can appear inevitable. True belief is a good thing for people, delusions and spells are bad; hence, you *help* those whose doctrines you assail, by leading them (or forcing them?) to a better cognitive state; they will make better decisions as a result of their enlightenment; they will be liberated by the unpleasant medicine you administer.

To see how and why one might escape this picture, it is useful to start with the evolutionary psychology Dennett (2007) offers. His speculations can be contrasted with different conjectures, with hypotheses that don't see the emergence of religious life in terms of cognitive defects. On Dennett's account, the predominance of religion in human societies stems from a prevalent psychological disposition, once adaptive, to think in ways that generate myths. Another possibility would suppose that religions played a role in the social lives of our Paleolithic ancestors, that they encouraged compliance with the norms of the small groups, that they fostered various types of solidarity, that they gave point and direction to individual lives. These ideas about social and psychological advantages accord with anthropological evidence about the status of gods as "guardians of morality," and also with some of the features sociologists discern when they study how people are attracted to particular religions. It would be wrong, however, to claim that my alternative is significantly better grounded than Dennett's: *both* are speculative. The important point, for present purposes, lies in their contrasting explanations of the predominance of religions. Where Dennett discerns an unfortunate cognitive glitch, I point to social and individual needs, some of them still important to people, which religions have met: societies that failed to develop religions were less good at meeting these needs, and, in consequence, they were less likely to transmit their forms of culture to descendant groups.

If you start with the thought that the predominance of religion in human societies is to be explained by a cognitive deficiency, you will tend to see your campaign for the eradication of myths in terms of a return to intellectual health. However difficult it may be for them to abandon their religions, people will be better off by repudiating the false beliefs that have held them captive. By contrast, if you suppose that the social factors toward which I have gestured have played a nontrivial role in the spread of the world's religions, you will wonder if there are psychological and social needs that the simple abandonment of religion will leave unfulfilled. You may even begin to wonder if the bracing tonics militant modern atheists conceive themselves as administering are – considered overall – a good thing.

My introduction of the orientation model is a vehicle for exploring these issues. Orientations are of primary importance in human lives, and people who cannot arrive at a reflectively stable set of values and aspirations rightly feel incomplete. Thoroughly secular people can have an orientation, as I have already conceded, but it is a fallacy to think that, for any religious person who currently fits the orientation model, that person can attain a cognitively superior orientation by rejecting the beliefs militant modern atheists discern as false. The cognitive gains can simply be outweighed by other forms of psychological and social loss.

To recognize this is not to patronize religious people, to view them as insufficiently astute or mature or courageous to come to terms with hard truths. Rather, it is to appreciate the difficulties contemporary social environments pose for the attainment of satisfying orientations outside of religious life. Dawkins' reflections on his own orientation reveal the ways in which humanly important issues can simply disappear from view:

> There is more than just grandeur in this view of life, bleak and cold though it can seem from under the security blanket of ignorance. There is deep refreshment to be had from standing up and facing straight into the strong keen wind of understanding:… (2003; quote in 2006, 397)
>
> The truths of evolution, along with many other scientific truths, are so engrossingly fascinating and beautiful; how truly tragic to die having missed out on all that! (2006, 320)
>
> But couldn't we also teach science as something to read and rejoice in, like learning how to listen to music rather than slaving over five-finger exercises in order to play? (1998, 36)

There is much to agree with in these passages, but they seduce readers – and Dawkins and Dennett too, I suspect – into thinking that *anyone* can orient a worthwhile life, one that will survive reflective probing, on the basis of contemplation of the cosmos as the sciences have revealed it.

Aristotle's great reflection on the good life in the *Nicomachean Ethics* closes with a picture of this kind, one that celebrates the life of informed contemplation. Yet, in the preceding books, Aristotle has pointed to other features that many people regard as central to the worth of the lives they lead: activity, contributions to social life, virtuous conduct, friendship, and so forth. It would be presumptuous for me to fathom the orientations of Dawkins and Dennett in any detail, but I am pretty sure that some of these alternative factors matter to them, too. They belong to a community of distinguished scholars, and to a subcommunity directed toward what they take to be an extremely important goal, that of spreading enlightenment as broadly as they can, and I strongly suspect that their membership in these communities and their sense of the contributions they make are an integral

component of their successful and satisfying orientations. They are probably in the best positions to articulate their aspirations and values, but I strongly suspect that any complete account of their sense of the worth of their lives would not consist in the passive pleasures of understanding various natural phenomena.

Even in affluent societies where citizens can be relatively well educated, the vast majority will never be able to recognize themselves as important participants in any impressive joint enterprise that contributes to knowledge and enlightenment. For large numbers of people, daily struggles to cope with threats to their physical well-being leave little opportunity for contemplation. Yet some institutions sometimes supply them with satisfying orientations, enabling them to rear their children with devoted love, helping them to create a less harsh and more just world for a few people around them. Some of these institutions are, of course, churches and mosques and synagogues.

Other things equal, it is preferable for these people to be mythically self-conscious, to enjoy the cognitive benefits toward which the militant modern atheists point, while also using the framework supplied by a religious tradition to direct their actions, in conjunction with others, toward valuable ends. But things are not always equal. Under some circumstances, the only psychologically and socially available ways of supporting a life that has any sense of worthy goals at which it aims, or any capacity for working with others to attain those goals, involve participating in traditions that cloud the messages Dawkins and Dennett want to deliver. Secular thinkers can regret that fact, but they should see it as a stimulus, not to break spells and abolish delusions, come what may, but to work toward an intellectually articulated and socially realized version of secular humanism that will permit satisfying orientations for the many people whose opportunities are currently limited.

Within the actual social environments in which contemporary people grow up, doctrinal entanglement can be expected to persist, not because the arguments directed against the doctrines are incomplete or because the people who hang on to belief in transcendent entities are too stubborn or too stupid, but because enlightened secularism has not yet succeeded in finding surrogates for institutions and ideas that religious traditions have honed over centuries or millennia. Until those surrogates are widely available, we need respect and tolerance for the doctrinally entangled. True enough, it would be better if their religion evolved to a state of mythical self-consciousness, but the costs for them – and sometimes for important social causes – that would attend the simple removal of false belief outweigh the benefits. Only those who approach these issues with the conviction that these matters are, from beginning to end, purely epistemic, who suppose that the belief model fits all religious lives, who think an evolutionary account will show how we've been HADD, will insist on breaking spells and ending delusions come what may.

If the doctrinally entangled should be considered in this way, the doctrinally indefinite are surely even more worthy of secular tolerance. For their errors are, at worst, those of their doctrinally entangled cousins. Moreover, once you recognize the importance that interpretation *must* play in reading religious texts, and once you appreciate the richness of many passages in these texts, it's easy to appreciate the predicament of the honest religious person who views parts of scripture as valuable for orienting her life and conduct, but who admits an inability to formulate any definite sentences that could articulate the content of the resonant words. There is something in those words that inspires and moves her, and she believes that they record some important past experiences – but the exact content of the inspiration and the precise character of the experiences is, she concedes, beyond her powers of definite expression.

I close by recapitulating two points I made earlier. First, commitment to any definite doctrine, mythical self-consciousness, or doctrinal indefiniteness must not incline the believer to slide into accepting normative claims that would underwrite conduct affecting the lives of others on the grounds that they express the will of some transcendent being – and I incline to think that a similar point holds even when her actions only affect herself. The orientation model insists that values come first. When further claims about what is to be sought or what is to be done are derived from religious doctrine, seen as expressing the commands of the deity, the believer has drifted from the orientation model into the belief model. Those to whom the belief model applies are, I have maintained, vulnerable to cogent critique. Hence, there must be a rigorous commitment to the priority of those values that can be shared with proponents of other religions and those who have no religion at all. Public reason must be thoroughly secular. The most that can be done is to offer particularly vivid and inspiring formulations of the common values.

My second agreement with militant modern atheism is almost a corollary of this. If one of the values we should share is that of self-determination, Mill's "pursuing one's own good in one's own way," then the practice of confining the thoughts of the young to the religious doctrines of parents – or to the secular attitudes of parents – is effectively the derivation of a specific value from the parental acceptance of doctrine. Enlightened parents should view the orientation model as suitable for their children, and should thus be concerned to open their children to a range of possibilities, while simultaneously helping them to acquire habits of thought that will equip them to choose wisely for themselves. Yet, until the environments in which many young people grow up have been radically transformed and enriched, involvement in a religious institution may protect the children, as it protected their parents, from the severe diminution of opportunities that our callous societies so frequently allow.

Acknowledgments

This paper was originally delivered at the Annual Meeting of the Society for Applied Philosophy, March 19, 2010. I am grateful to Taylor Carman and Wayne Proudfoot for valuable comments on an earlier draft, to the audience at the SAP meeting for some excellent questions, and to Richard Harries for a stimulating conversation before the lecture.

References

Armstrong, K. (2009). *The Case for God.* London: Bodley Head.
Dawkins, R. (1998). *Unweaving the Rainbow.* Boston: Houghton Mifflin.
Dawkins, R. (2003). *A Devil's Chaplain.* Boston: Houghton Mifflin.
Dawkins, R. (2006). *The God Delusion.* Boston: Houghton Mifflin.
Dennett, D. (2007). *Breaking the Spell.* New York: Penguin.
Dewey, J. (1933). *A Common Faith.* New Haven: Yale University Press.
Kant, I. (1785/1981). *Groundwork of the Metaphysics of Morals.* Indianapolis: Hackett.
Mill, J.S. (1859/2008). *On Liberty.* Oxford: World's Classics.
Russell, B. (1967) *Why I am not a Christian.* New York: Simon & Schuster.

31

Why Naturalism Cannot Account for Natural Human Rights

NICHOLAS WOLTERSTORFF

The concept of natural rights, and in particular of natural *human* rights, was first employed systematically by European canon lawyers in the 12th century.[1] Gradually, it became a prominent component within the moral and political thought of Western Europeans. The US Declaration of Independence ringingly declared, "We hold these truths to be self-evident, that all men are…endowed by their Creator with certain inalienable rights, that among these are life, liberty, and the pursuit of happiness." In the 20th century, there were movements, in the United States and various European countries, for women' rights, children's rights, the rights of labor, and civil rights. With the adoption by the United Nations in 1948 of the Universal Declaration of Human Rights, the concept of natural human rights began to play a prominent role within the moral and political thought of human beings around the globe.

What must be added is that not everybody in the modern world accepts and employs the concept of natural human rights. Classical utilitarians do not; one's choice of what to do is to be guided solely by comparative estimates of life-goods to be brought about. Some Christians do not; we are always and only to act out of gratuitous benevolence, not out of respect for rights. A good many Muslims dismiss the idea of natural human rights as an invention of the West. And Karl Marx had no use for the idea, as is clear from such passages as the following: "None of the so-called rights of man goes beyond egoistic man, man as he is in civil society, namely, an individual withdrawn behind his private interests and whims and separated from the community" (McLellan 1977, 54) and: "The freedom in question is that of a man treated as an isolated monad and withdrawn into himself…The right of man to property is the right to enjoy his possessions and dispose of the same arbitrarily, without regard for other men, independently from society, the right of selfishness" (McLellan 1977, 53).

In the *Weekly Standard* of March 25, 2013 there was a report by Andrew Ferguson, one of the senior editors of the *Standard*, of a workshop that took place in October 2012 in the Berkshires titled "Moving Naturalism Forward." The participants in the workshop were

1 I base this claim on the groundbreaking work of Brian Tierney (1997).

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

self-proclaimed naturalists of the physicalist sort. The topic of the conference was how the cause of physicalism could be advanced.

Ferguson reports the participants as all accepting and employing the distinction between the manifest image of the world and the way things really are. The manifest image of the world is how things appear to us; it includes colors and sounds, pains and itches, tastes and smells, desires and choices. But science, so the participants in the workshop claimed, tells us that this manifest image is entirely illusory. "Science, in the words of one of the participants, the philosopher Daniel Dennett, is a 'universal corrosive,' destroying illusions all the way up and all the way down, dismantling our feelings of freedom and separate selfhood, our morals and beliefs, a mother's love and a patient's prayer. All in reality are just 'molecules in motion'" (Ferguson 2013).

Ferguson reports Dennett as complaining that there were still a few philosophers – very few! – who stubbornly refused to incorporate the conclusions of science into their philosophizing, continuing to play around with outmoded ideas like morality and sometimes even the soul. The one who drew universal scorn and contempt at the workshop was the eminent American philosopher, Thomas Nagel, whose recent book, *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*, infuriated the participants. "I am just appalled," said Dennett, "to see how in spite of what I think is the progress we've made in the last 25 years, there's this sort of retrograde gang. They're going back to old-fashioned armchair philosophy with relish and eagerness. It's sickening. And they lure in other people. And their work isn't worth anything – it's cute and it's clever and it's not worth a damn" (Ferguson 2013).

It is noteworthy that, in that last sentence, Dennett employs the concept of *worth*; his comment, that the work of the "retrograde gang" of non-naturalists is worthless, carries the clear implication that his own work and that of his fellow naturalists has worth. But the term "worth" is not to be found in the vocabulary of physics. Has Dennett inadvertently thrown his lot in with the retrograde gang?

Ferguson does not report Dennett as having been pressed on this point. Had he been pressed, he would no doubt have appealed to a distinction commonly made by present-day naturalists, between two kinds of language. Parallel to the distinction between the manifest image of the world and the way things really are is the distinction between folk language and the language of physics. Dennett was using folk language when he implied that his own work and that of his fellow physicalists had worth. Folk language is indispensable; there is no way of translating what we say in folk language into the language of physics. But we must not suppose that when we employ folk language, we are describing things as they really are. Only when we use the language of physics are we doing that.

The only language Dennett ever uses is folk language; he is not himself a physicist. I would not be surprised to learn that in his use of folk language, he freely speaks not only of worth but also of rights; neither would I be surprised to learn that he has on occasion described what he was doing as demanding his rights, or as standing up for the rights of others. But though in his usage of folk language he may well use the term "rights," it is nonetheless his view that in reality there are no rights. The vocabulary of physics does not include the term "rights."

I find the suggestion that one can speak with the vulgar while thinking with the physicists full of conundra. On this occasion, I will forego identifying those conundra. What I will do, instead, is rehearse an argument about rights that I have developed in some of my writings. Taking for granted the widely shared assumption that there are natural human rights, I have argued that whereas theism, in certain of its versions, has the resources for

giving an account of natural human rights, there is, to the best of my knowledge, no plausible nontheistic account of such rights. I will not argue that there could not possibly be such an account, only that there is none at present.

Dennett and his cohorts would, of course, disagree with the assumption that there are rights. Though they use rights-talk, they hold that in reality there are no rights. But not only would they disagree with my assumption, they would resist my argument. To accept my argument would mean that in their use of folk language, they would have to use God-talk when accounting for rights. It is clear from their writings, however, that they have resolutely purged from their folk language all talk of the God of Judaism, Christianity, and Islam. Why they are resolved to avoid God-talk while freely speaking of worth, of rights, and of a mother's love, given their conviction that God, worth, rights, and love are equally nonexistent, is a nice question. If it is OK to use love-talk while holding that there is no such thing as love, why not use God-talk while holding that there is no God?

The term "naturalism" is commonly used not only for the physicalism espoused by the participants in the Berkshire workshop, but for other, more capacious, ideologies as well. Some, for example, have identified Aristotle as a naturalist. Not only was Aristotle not a physicalist, but his metaphysical scheme included a god whose function in the scheme was to think about itself. I have no idea whether Dennett and his cohorts would be willing to include, in their idiolect of folk language, talk about a god of the Aristotelian sort. I am confident in asserting, as I just did, that they would not be willing to include talk about the theistic God of Judaism, Christianity, and Islam.

What Are Natural Human Rights?

A right is always a right *to* something; and always, when one has a right to something, one has a right to being treated a certain way by one's fellows or by oneself. We do speak of rights to things other than ways of being treated; we speak, for example, of a right to a seat on the train. But if I have a right to a seat on the train, then I have a right to being assigned a seat by the relevant official. And so it is in general: always, when one has a right, one has a right to being treated a certain way.

One has a right to being treated a certain way only if being treated that way would be a good in one's life. Rights are never to life-evils, only to life-goods. However, it is only to some of those ways of being treated that would be a good in one's life that one has a right. A whimsical example that I have on occasion given is this: It would be a good in my life, indeed, a *great* good, if the Rijksmuseum in Amsterdam were to give me Rembrandt's *The Jewish Bride* to hang on my living room wall, along with a paid security force to stand guard. But I do not have a right to their doing that; they are not wronging me by not doing that. One of the principal challenges facing anyone who wishes to develop an account of rights is to explain what it is that gives one a right to certain ways of being treated that would be goods in one's life but not to others.

One has a right to being treated a certain way if being treated that way is *due* one – alternatively, if one *has a legitimate claim* to being treated that way. These are truisms; they cast some light on the concept of a right, but not much. We want to know what accounts for the fact that some ways of being treated are *due* one whereas others are not. What accounts for the fact that one has a *legitimate claim* to some ways of being treated that would be a good in one's life, whereas to others one does not?

The most common view on this matter, a view that I share, is that rights are grounded in the worth, the dignity, the excellence, the estimability of the rights-holder: one has a right to someone treating one a certain way just in case, if they did not treat one that way, they would not treat one as befits one's worth. Rights are what respect for worth requires.[2]

A good many of the rights that we possess have been generated in us, or bestowed on us, by an action of some human being or organization. Michael promised me that tomorrow he would return the book I loaned him; his making that promise generated my right to his returning the book tomorrow. I have a right to a monthly Social Security check from the US government because the US Social Security legislation bestowed that right on me. But not all of the rights we possess have been generated in us or bestowed on us by human action; and some of those that have been so generated or bestowed are ones that we would have possessed anyway. Such rights are *natural* rights. The right not to be murdered is an example.

The concept of a *human* right proves more complicated.[3] We can get hold of the concept of a human right either by starting from a list of what are generally cited as human rights and trying to surmise the concept implicit in the list, or by starting from the common explanation of the concept. It turns out that these two approaches yield significantly different results.

Suppose that one looks over the lists of rights found in the various UN declarations on human rights. One will be confronted by the appearance on the lists of items that do not fit the concept as commonly explained. The common explanation, shortly to be unpacked and refined, is that a human right is a right that one has just by virtue of being a human being. Article 23 of the Universal Declaration of Human Rights says, "Everyone has the right to work, to free choice of employment, to just and favourable conditions of work, and to protection against unemployment." To possess this right, it is not sufficient that one be a human being; one must be a certain *kind* of human being, namely, one capable of working. Not all human beings are capable of working; infants are not, those sunk into advanced dementia are not. Such human beings are not wronged by not having work.

The explanation of this anomalous situation seems to me clear. Those who have written about human rights, including those who composed the UN documents, typically have not had in mind all human beings whatsoever. They have not had in mind human beings in advanced dementia, those in a permanent coma, or infants. They have had in mind those human beings who possess the traits and capacities that make a human being a *person*; they have had in mind *full-fledged human persons*. Human beings in advanced dementia or in a permanent coma are not full-fledged human persons. They once were; but they are that no more. They remain human beings, however.

Given the discrepancy to which I have pointed, we have to make a choice. Shall we be guided by the standard explanation of a human right and understand such a right as a right that one has just *qua* human being, or shall we be guided by the common lists of human rights and understand a human right as a right that one has *qua* human *person*?

I think that those of us who are philosophers should, after getting clear on the distinction, explore both the grounding of *human* rights and the grounding of what one might call *human-person* rights. When I have argued that there is no plausible nontheistic grounding

2 I develop this account of rights much more fully in Wolterstorff (2008).

3 Most of what follows in the remainder of this chapter follows closely my discussion in Chapter 7, "On Secular and Theistic Groundings of Human Rights," in Wolterstorff (2012).

of human rights, it was *human* rights that I had in mind, not *human-person* rights. I think it is possible to give an adequate and plausible nontheistic grounding of human-person rights;[4] it is only concerning human rights that I make my claim. In what follows I will, accordingly, be speaking only of those rights that one has just *qua* human being.

To possess a particular right, one must have a certain *status*, as I shall call it. For example, to have a right to the prize designated for the winner of the Boston marathon, one must have the status of coming in first and not having violated any rules of the contest. A human right is a right such that the status sufficient for possessing the right is that of being a human being. One does not have to be a *Greek* human being, a *male* human being, or an *educated* human being. One does not have to be any particular kind of human being whatsoever, not even a human being capable of functioning as a person. It's enough that one have the status of being a human being.

The rights commonly cited as human rights have to be understood as *prima facie* rights; in a given case, one's *prima facie* human right to the good of being treated a certain way may be outweighed by a conflicting *prima facie* right of someone else. Given the trumping force of rights, it can only be outweighed by other *prima facie* rights, however, not by goods to which no one has a *prima facie* right. It may be that certain *prima facie* human rights are never outweighed by any other *prima facie* rights whatsoever. I judge that one's right not to be physically mutilated is an example of such; one's right not to be physically mutilated for the pleasure of the mutilator is even more obviously an example.

My right to be treated a certain way may be a human right even though I find myself in a circumstance where I do not have that right – that is, in a circumstance where it is not even a *prima facie* right of mine. That would be the case if I found myself in a circumstance where it was impossible to be treated in the way specified; my possession of the right would in that case be *blocked*. Human rights are not peculiar in this regard; it's true of rights in general that one does not have a right to be treated a certain way if it is impossible that one be treated that way.

Given that human rights can be blocked by straitened circumstances, they are not to be equated with the universal rights of human beings – understanding by a "universal right," a right that necessarily every human being always has. Some are universal – once again, the right not to be physically mutilated for the pleasure of the mutilator. But others are not; for example, the right to fair and reasonable access to means of sustenance. What can be said is that if some right is a human right, then anyone in a circumstance where that right can be honored possesses that right.

As with many distinctions, the distinction employed in the preceding paragraphs, between one's *status* and one's *circumstance*, is such that sometimes it's not clear whether a description is a description of one's status or a description of one's circumstance. But we need the distinction. If we said that a human right was a right such that sufficient for having it is that one be a human being, it would turn out that there were very few human rights: only those that are universal. I hold that among our human rights is the right to fair and reasonable access to adequate means of sustenance. But if we insisted that a right, to be a human right, has to be universal, this would not be a human right, since there are circumstances in which that right cannot be honored – extreme and pervasive drought, for example.

4 I have tried my hand at doing so in Chapter 8, "Grounding the Rights We Have as Human Persons," in Wolterstorff (2012).

To repeat: a human right is a right such that the status sufficient for possessing it is that one be a human being. Is this also a *necessary* condition? If animals have rights, then, since humans are animals of a sort, it's likely that among our rights are some that other animals possess as well. So suppose that a certain right is such that not only is the status of being human sufficient for possessing the right, but the status of being a non-human higher primate is also sufficient. Consider, yet again, the right not to be mutilated for the pleasure of the mutilator. I judge that not only human beings but also other higher primates have this right – and other animals as well. Should we refrain from calling this right a *human* right on the grounds that non-human animals also possess it?

It would represent a sharp break with the common usage of the term "human right" if we did say that a right is not a human right if some animals also have it. Some human rights are rights that animals also have; it is not a necessary condition of a right's being a human right that its holder be a human being.

A common assumption in the literature on human rights is that *not all* human rights are shared with animals. This assumption seems to me correct. I judge that all human beings have a right to a degree and extent of medical and nursing care aimed at preserving life that no animal has a right to – difficult though it is to specify just what that degree and extent is. This right is uniquely human. And if it is uniquely human, then the entire package of human rights is not shared with any non-human animal. The package is uniquely human.

On Grounding Human Rights

The UN declarations all affirm or assume that human rights accrue to human beings on account of the dignity, the worth, the excellence, the estimability that human beings possess. Almost everyone who has written on the topic shares this assumption.[5] I share it as well.

Worth, dignity, and excellence do not settle on things willy-nilly; always, there is something about the thing that gives it worth, something that accounts for its worth, something on which its worth supervenes. If the piano sonata that you composed is a fine sonata, then there is something about it that makes it fine. We may find it difficult, perhaps even impossible, to put into words what that is; but it makes no sense to say that it's a fine sonata but that there is nothing about it that makes it fine.

The big challenge facing the theorist who believes that human rights are grounded in the worth of the rights-bearer is to identify that feature of human beings (or combination of features) on which that worth supervenes. To successfully identify that feature would be to offer a dignity-based *grounding* of human rights. A striking feature of the UN documents is that though they all affirm or assume that all human beings have dignity and that their human rights are grounded in that dignity, they refrain from making any attempt to account for that dignity.

Let's be clear on the sort of feature we are looking for. First, it has to be a feature that is ineradicable from human beings, a feature that no human being can lack so long as he or she exists. The reason for this goes as follows. Human rights are such that the status sufficient for having them is that of being human. So if human rights are grounded in dignity, then the dignity that grounds those rights must be ineradicable from the status of being

5 I know of only one exception, namely, Alan Gewirth. In Wolterstorff (2008, 335–340), I present and critique Gewirth's theory.

human; otherwise, the status sufficient for having those rights would not be that of being human. And if that dignity is ineradicable from the status of being human, then the feature(s) on which that dignity supervenes must likewise be ineradicable from the status of being human; otherwise, one could have the status without having that feature (those features) and hence without having that dignity. But the status of being human is ineradicable from human beings. So the feature we are looking for must be ineradicable. Infants, Alzheimer's patients, those in a permanent coma, anyone who is a human being must possess the feature we are looking for.

Second, given that the entire package of human rights is not shared with any of the non-human animals, the feature (or combination of features) on which the dignity that grounds human rights supervenes has to be one that no non-human animals possess. It must be a uniquely human feature.

In discussions of dignity-based attempts to ground human rights, one finds it commonly assumed that the feature we are looking for has two additional properties. It is assumed that the feature gives to each human being a worth greater than that possessed by any non-human animal, an animal-transcending worth. And it is assumed that the worth human beings have on account of possessing the feature is equal. Human beings vary greatly in the worth they have on account of native endowments, achievements, and so forth. But it is assumed that there is something about our being human that gives us equal dignity. *Qua* human, we are equal in dignity.

It is not required by the concept of a human right as explained in this chapter that the feature we are looking for has these two properties: that of giving us animal-transcending dignity and that of giving us equal dignity. All that is required is that it be ineradicable from human beings and that it be uniquely human. Though the concept of a human right does not require that the feature we are looking for has the properties of giving us animal-transcending dignity and of giving us equal dignity, in the discussion that follows I will nonetheless take note of whether or not the various features proposed, as accounting for the worth that grounds human rights, do have these additional properties.

There is another condition for success in developing a dignity-based grounding of human rights that is of a sort quite different from finding a feature of human beings that is ineradicable and uniquely human. I and those other theorists who engage in the project of trying to ground human rights in human dignity do so because we have the intuitive sense that every human being has dignity just by virtue of being human, and that it is because of this dignity that certain ways of treating that human being are morally unacceptable; to treat her thus would be to violate her dignity. The example I have been using is that of mutilating a human being for the pleasure it gives the mutilator.

Given that we come to the project of grounding human rights with this intuition, a condition of success is that we find a feature that makes our intuition intelligible. We want to arrive at the point where we can say, "Yes, now I understand why even a person in a permanent coma has a right not to be mutilated for the pleasure of the mutilator, why even she has a dignity that makes that impermissible." With only a modicum of imagination, we can come up with features that are ineradicable from human beings and uniquely human; having the human genome (whatever that may be) as one's genetic makeup would be an example. But most of those will be such that, should someone suggest that the dignity that grounds human rights supervenes on one of those features, our response will be bafflement. The suggestion does not illuminate our prior intuition. "What does that have to do with dignity?"

Does the Capacity for Rational Agency Ground the Dignity that Accounts for Natural Human Rights?

Almost all secular, nontheistic attempts to account for human rights that are to be found in the literature are dignity-based accounts; and almost all of those, in turn, are what one might call *capacity accounts*; they hold that the worth that accounts for human rights supervenes on a certain capacity that human beings have.[6]

There is remarkably little variation in the capacity proposed. To the best of my knowledge, it is always either the capacity for rational agency in general or some specific form of that capacity, such as the capacity for acting out of duty, the capacity for acting on an apprehension of the good, or the capacity for forming, implementing, and revising a plan of life. The basic idea goes back to Immanuel Kant. I will conduct my discussion in terms of the capacity for rational agency in general; what I say about that capacity will apply to the specific forms mentioned.

I share the conviction that the capacity for rational agency gives to those who possess it great worth. To be able to act for reasons is extraordinary; and wherever we find a creature that possesses this capacity, we should treasure that creature and treat her with the respect due her as someone who possesses it. The questions before us here are whether this capacity is ineradicable from human beings, whether it is uniquely human, and whether noticing that human beings have this capacity makes intelligible our intuitions about the dignity and rights of each and every human being.

As a preliminary matter, notice that the capacity for rational agency comes in degrees, with the consequence that if the dignity that accounts for human rights supervenes on having the capacity for rational agency, then that dignity also comes in degrees. Thus, the proposal that the dignity that accounts for human rights supervenes on the capacity for rational agency violates the common equal-worth assumption. But, as we have seen, it is not required of a dignity-based grounding of human rights that it satisfy the equal-worth assumption. So let's move on.

A point that is relevant is that the capacity for rational agency appears not to be unique to human beings; porpoises and chimpanzees, for example, seem also to have it. Thus, this proposal fails to satisfy the requirement that the feature(s) we are looking for be uniquely human.

If it were the case that the highest degree of rational agency possessed by non-human primates was lower than the lowest degree possessed by any human being, then we could say that the dignity that accounts for human rights supervenes on possessing the capacity for rational agency to a degree not found in any non-human animal. But that is clearly not the case. Some human beings are less capable of engaging in rational agency than are some mature, well-formed, non-human animals. Moreover, and more importantly, some human beings do not have the capacity for rational agency to any degree: newborn infants, for example, and those sunk deep into dementia. The capacity for rational agency is not ineradicable.

Suppose that rather than immediately abandoning the rational-capacity approach because some human beings do not have the capacity, we instead look for a different relation to the capacity than that of actually possessing it. What might that different relation be?

Let me prepare the ground for a suggestion by first asking whether infants bear some relation to the capacity, other than possessing it, which illuminates our intuition that they

6 An exception is Ronald Dworkin's account, which I present and critique in Wolterstorff (2008, 333–334).

too possess the dignity that grounds human rights. I think they do. What illuminates our intuition in their case is noticing that most of them bear the *promise* of having the capacity for rational agency. If one prizes the capacity for rational agency and ascribes dignity to those who have the capacity, how could one not ascribe dignity to those who bear the promise of having that capacity? Of course, it is not true in general that the worth one has on account of bearing the promise of being in some worth-bestowing condition is the same as the worth one has on account of actually being in that condition; the honor due the heir apparent to the Prince of Wales is quite different from that due the Prince of Wales.

Those sunk deep into dementia neither have the capacity for rational agency nor bear the promise of having it; the same is true for those suffering from severe and irreversible brain injury and those in a permanent coma. But do they perhaps stand in yet another relation to the capacity for rational agency, a relation that illuminates our intuition that they too possess the dignity that grounds the right not to be mutilated for the pleasure of the mutilator?

I think they do. What illuminates our intuition in their case is noticing that they once had the capacity for rational agency. They carry the history of having been capable of rational agency; each of them is, as it were, a relic of having had it. That is part, at least, of why we regard them as having the dignity that grounds rights.

Recall, once again, what we are looking for. We are looking for a feature that is unique to human beings, that is ineradicable, and that illuminates our intuitions concerning dignity and rights. We first observed that *having* the capacity for rational agency does not fill the bill. Having the capacity is neither unique to human beings nor ineradicable; infants and those sunk into dementia do not have the capacity. We then observed that most infants bear the promise of having the capacity; that illuminates why we ascribe dignity to them. And we observed that those sunk into dementia are living relics of having the capacity; that illuminates why we ascribe dignity to them. The feature suggested by the steps we have gone through is the disjunctive property of *having the capacity for rational agency or holding the promise of having it or once having had it*.

Is this the feature that we have been looking for? Does this disjunctive property fill the bill? It does not, for the same reason that *having* the capacity for rational agency does not fill the bill. Possessing that disjunctive property is not unique to human beings; some non-human animals possess it. And some human beings do not possess it; it is not ineradicable from human beings. There are severely impaired human beings who do not now have the capacity for rational agency and never did and never will. Our search remains unsuccessful.

There is also a question as to whether the illumination condition has been satisfied for those human beings who do possess that disjunctive property. Our intuition that well-formed mature human beings have the right not to be mutilated for the pleasure of the mutilator is illuminated by our noticing that they have the truly remarkable capacity of rational agency. Our intuition that infants have that right is illuminated by our noticing that they bear the promise of having that truly remarkable capacity. Our intuition that those sunk deep into dementia have that right is illuminated by our noticing that they once had that truly remarkable capacity. The question to consider is whether, for a human being of any of these sorts, our intuition that he has that right is illuminated by noticing that he has the property of *having the capacity for rational agency or holding the promise of having it or once having had it*.

I doubt it. Our intuition is illuminated in the case of an infant by noticing that he bears the promise of having the capacity for rational agency; I doubt that it is illuminated by noticing that he possesses the property of having or holding the promise of having or once having had the capacity for rational agency.

Does Personhood Ground the Dignity that Accounts for Human Rights?

Being a person includes more than the capacity for rational agency. To be a person is also to be a center of trust and mistrust, of hope and despair, of love and hate; a center of emotions, of feelings, of beliefs, of intentions, of sensory perceptions, of plans, of private reveries.[7] It is to be a "center" of all such states and actions in the sense that we do what Kant describes as "attaching the 'I' to our representations." That is to say: we ascribe a certain hope to ourselves, a certain regret to ourselves, a certain sensation to ourselves, a certain belief to ourselves, a certain action to ourselves, and so forth. "*I* hope that tomorrow will bring better weather." "*I* knocked over the vase."

To possess this capacity for self-ascription, along with whatever else belongs to being a human person, is truly remarkable. Personhood is to be treasured wherever we find it; those who possess it have great dignity. We have no good reason for thinking that any non-human animal can function as a person. Further, though I agree that our intuition that mature, properly functioning human beings have a dignity that grounds human rights is illuminated by noticing that they have the capacity for rational agency, I think that even more light is cast on that intuition by noticing that they are persons. It's a violation of their personhood to mutilate them just for the fun of it.

The question before us, however, is whether *being a person* is the feature we are looking for. Does it fill the bill? Does the dignity that grounds human rights supervene on our personhood? It does not. Most of the problems that arose for the rational agency proposal also arise for the personhood proposal. Tiny infants cannot yet function as persons; those sunk deep into dementia can no longer function as persons.

Newborn infants do, however, bear the promise of functioning as persons; and those sunk deep into dementia and those in a permanent coma were once capable of functioning as persons. So might it be that the feature we are looking for is this one: the property of *functioning as a person or holding the promise of so functioning or having once so functioned*? Does this property fill the bill?

It does satisfy the uniqueness condition; no non-human animal has this property. In that respect, the personhood proposal for the grounding of human rights is superior to the rational agency proposal. It does not, however, satisfy the ineradicability condition. Sad to say, some human beings never function as persons, so they do not have this property. It's possible to be a human being and lack the property.

A Final Attempt at Secular Grounding

Notice that no matter how severely impaired a human being may be, she retains human nature. Indeed, the idea of an *impaired* human being presupposes the idea of human nature. It's by virtue of possessing human nature that a human being can be impaired; that is, malformed and not capable of functioning properly. Those who have lost all capacity for rational agency are not just statistically unusual – something has gone wrong with them.

7 In Chapter 8, "Grounding the Rights We Have as Human Persons," of Wolterstorff (2012), I develop the richer account of what it is to be a person that I am here pointing towards.

Human nature is remarkable. Human nature is such that mature and properly formed exemplars of this nature function as persons and have the capacity for rational agency. No non-human animal has so remarkable a nature. So might it be that noticing that even the most impaired human beings still possess human nature illuminates our intuition that even they have the dignity that grounds rights?

Suppose one is holding such a human being in one's arms and finds oneself convinced that she has a dignity that makes certain ways of treating her just wrong. Suppose one is at a loss, however, to explain why one believes this. But then someone remarks that she too has human nature, and expatiates on how noble that nature is. Would that illuminate one's intuition? Possessing human nature is obviously ineradicable from human beings and is uniquely human; and if it gives human beings worth, it gives them equal worth. But does it illuminate our intuition? Of course, there are a good many writers nowadays who reject the idea of human nature; for them, this proposal is a nonstarter.

How can we wrap our minds around this proposal? Let me make a suggestion. Think of our human nature as our design plan, and then consider what we would say about other cases of well-formed and malformed exemplifications of a design plan. Suppose that my two neighbors own the same model of automobile; a certain model of Jaguar, let's say. The example that my neighbor to the east owns is well formed in all respects and I admire it enormously. The example that my neighbor to the west owns was in a wreck and was towed into his garage; the mechanics and body repair shops all tell him that to repair it would require such extensive replacement of parts that it's best to scrap it.

Would I advise my neighbor to the west to reject this advice and instead to treasure his automobile as something of great worth? Would I tell him that even though it's a wreck, nonetheless it is truly admirable on account of its design plan and that he should keep it under a dust cover and every now and then lift the cover to admire it? I would not. I would advise him to sell it for scrap. It has no worth other than its worth as scrap metal, plastic, leather, and glass.

The point of the analogy is obvious: noticing that even the most impaired human beings possess human nature, and reflecting on the nobility of that nature, does not illuminate our intuition that these human beings have the dignity that grounds human rights.

Does the *Imago Dei* Ground the Dignity that Accounts for Human Rights?

If even the personhood account fails, it's hard to see how any dignity-based account of human rights that appeals to capacities or to a way of functioning could be successful. And if the human nature account also fails, then it would seem that all that's left to consider is whether our standing in some relation to some being (or beings) gives us the requisite worth. Obviously, it would have to be a relation such that standing in it does not presuppose functioning as a person or possessing the capacities of rational agency.

Easily the most plausible candidate for the being in question is God, assuming that God exists. Not only is God the most plausible candidate, it's hard to think of any other candidate that has any plausibility whatsoever. So let's turn from the secular accounts we have been considering and ask whether theism, in some form, possesses the resources for grounding natural human rights that secular ideologies do not. Of course, a good many people in present-day Western society believe that God does not exist. Nonetheless, they too can consider whether theism, in some form, possesses the resources for grounding human rights.

What might be the relation of human beings to God that gives to each and every one of them the dignity that accounts for human rights? A suggestion commonly made by theologians is that it is the relation to God of bearing the image of God, the *imago dei*. To decide whether this suggestion is tenable, we obviously have to know what constitutes the image of God.

The history of Christian (and Jewish) thought contains a quite astounding diversity of proposals. To the best of my knowledge, these are all of one or the other of two sorts. Some writers understand the image of God to consist of resembling God with respect to the possession and exercise of certain capacities; others understand it as consisting of resembling or representing God with respect to exercising a certain role in creation – the role of "having dominion," for example, as the opening chapter of Genesis puts it, or the role of "being about God's business," as John Locke puts it.

All the roles proposed by those who favor the second understanding require exercising certain capacities. Thus, the main points made in previous sections of our discussion apply here as well: whatever functions of personhood or capacities of agency are required to possess the image of God, some human beings do not yet have those, some no longer have them, and some never have them. But if some human being does not have those functions or capacities, then she does not resemble God with respect to having them, nor does she resemble or represent God with respect to her role in creation. If the image of God is understood in either of these two traditional ways, then some human beings lack the image of God.

Though I know of no theologian who explicitly takes note of this implication of the traditional understandings of the *imago dei*, we can be confident that few if any would be happy with this implication. Though there is a long theological tradition which holds that the image of God has been impaired by sin, few theologians would accept that some human beings lack the image entirely. So is there some way of understanding the image of God that does not have this implication?

Our discussion in the immediately preceding section suggests such a way. No matter how impaired a human being may be, she nonetheless possesses human nature. And in possessing such a nature, she resembles God in a certain way. Human nature is like the divine nature in that mature, properly functioning specimens of human nature possess God-resembling functions and capacities; by exercising those, they can play a God-resembling or God-representing role in creation.

So suppose that in place of the standard capacity-resemblance and role-resemblance accounts of the image of God, we adopt what one might call a *nature-resemblance* account: the image of God consists in possessing a nature that is like the divine nature, in that mature, properly functioning specimens of this nature possess God-resembling functions and capacities that enable them to play a God-resembling or God-representing role in creation.

Every human being bears the image of God so understood. So the question before us is whether our bearing the *imago dei* so understood gives to us an ineradicable and uniquely human dignity. Does the dignity that accounts for natural human rights supervene on standing to God in the relation of bearing the image of God – image being understood along nature-resemblance lines?

We have already seen that merely possessing human nature does not give one the relevant dignity. So the question to consider is whether *resembling* God with respect to one's nature gives to one the relevant dignity. It does not – not so far as I can see. If possessing human nature does not itself give one the requisite dignity, why would resembling God with respect

to possessing such a nature do so? The fact that the design plan of the wrecked car in the garage of my neighbor to the west resembles the design plan of the well-functioning car of my neighbor to the east does not somehow give a worth to that wrecked car that otherwise it would not have.

A Theistic Account of Human Dignity

The question before us is whether, according to at least some version of Judaic, Christian, or Islamic theism, there is some relation in which each and every human being stands to God by virtue of which they all have the ineradicable and uniquely human dignity that grounds human rights. It would have to be a relation such that standing in it does not presuppose having the functions of personhood or the capacities of agency. We have just now seen that standing to God in the relation of bearing the image of God is not the relation we are looking for. I suggest that the relation we are looking for is that of being loved by God, love being understood as the desire for friendship.[8]

Imagine a good monarch who is loved by all his subjects; he bestows on all of them the great good of a just political order that serves the common good. But he is lonely. So, in addition to acting as a benefactor to all his subjects, he decides to select some of them as ones with whom he would like to be friends. This is an honor for the ones chosen. "I am honored that you would choose me to be your friend," they say. No doubt, over the course of time, various goods in their lives will ensue from the monarch's being friends with them. But just to be chosen as one with whom the monarch would like to be a friend is an honor. Those chosen will cite that fact under "Honors" in their resumes. Those not chosen will be envious.

Now for the crucial point. To be honored is to have worth bestowed on one. Admittedly, this is mysterious. But an indication of the fact that this is what happens is that to be honored is to acquire a new ground for respect and, hence, new ways in which one can be demeaned and wronged; and that's possible only if there has been some alteration in one's worth. The most obvious way in which one can now be demeaned and wronged is by having the honor itself belittled. One who has not been chosen by the monarch for friendship might sarcastically remark to someone who has been chosen, "Big deal!" The latter has a right to be angry: this is a snub. Depending on the situation, the snub might also count as a snubbing of the one who bestowed the honor; but in any case, it amounts to snubbing the one honored. Honoring bestows worth.

It's important to add that not every case in which a group thinks they are honoring someone by having an official utter some such words as "I hereby bestow on you the honor of so-and-so" is in fact a case of bestowing an honor on that person. The Ku Klux Klan may think they are honoring someone by having one of their officials pronounce those words over someone. But, in fact, no honor has been bestowed. If the person over whom those words were pronounced comes to his moral senses, he will not mention under "Honors" in his resume that he was honored by the KKK.

Sometimes an honor is bestowed as a way of recognizing worth, but not always. The monarch's choice of a few subjects as those with whom he would like to be friends is not his

8 In Wolterstorff (2008), I said that the love in question is love as attachment; I now think that that was a mistake.

way of declaring that these are the most estimable of his subjects. The monarch must spot some potential for friendship in those he chooses. In some cases, there will be obstacles of various sorts to the realization of this potential; the monarch and the ones chosen will then have to work at overcoming those obstacles. But that's different from the monarch's selecting certain people for friendship because they are the most estimable in the realm and as a way of declaring that they are.

The application of the analogy is obvious. Suppose that one is chosen by God as someone with whom God desires to be a friend. This is to be honored by God. And to be honored by God is to have worth bestowed on one. Add now that every human being has the honor of being chosen by God as someone with whom God desires to be friends and that this desire endures. Then every human being has the ineradicable and equal worth that being so honored bestows on one.

A question to consider is whether, on this view of things, it is purely whimsical and arbitrary of God to choose human beings as the creatures with whom God wants to be friends. Might God just as well have chosen crocodiles? No. I remarked that though the monarch does not look for the most estimable people in the realm when settling on those with whom he would like to be friends, he does look for those in whom he sees potential for friendship. Crocodiles lack the potential for being friends with God. Being friends with God is incompatible with crocodilian nature. To be a friend with God, one has to have the nature of a person. Crocodiles at their best cannot be persons. Of all the animals, it's only human animals that can function as persons and can thus be friends with God. The worth bestowed on us by God's desire to be friends with us is both uniquely human and animal-transcending.

The same consideration that makes it understandable why God did not choose crocodiles for friendship makes it understandable why God chose human beings. Since it's in our nature to be persons, we have the potential for friendship with God. Of course, there are blockages to the realization of that potential that have to be overcome by God and us. The moral breach between us of our having wronged God will have to be repaired; and those who cannot presently function as persons will have to be healed, in this life or the next, of that deep malformation. Though God's desire for friendship with some human being does not presuppose that that human being presently has the capacities necessary for the satisfaction of that desire, it does presuppose that that human being *will someday* have those capacities, in this life or the next.

Earlier, I said that, so far as I could see, possessing human nature does not, by itself, provide us with a worth sufficient for grounding human rights, and that resembling God with respect to possessing such a nature does not add the requisite worth. It now turns out that human nature and the *imago dei* are not irrelevant to a full theistic grounding of human dignity. They make it understandable that God would choose human beings for friendship and not the non-human animals. Given our nature, we have the potential for friendship; given their nature, they do not.

Notice that the answer I have given to the question whether God's desire for friendship with human beings and not with the other animals is whimsical and arbitrary is not an *explanation* of God's desire for friendship with human beings. Our possessing human nature provides the potential for friendship between God and us; it's a necessary condition for friendship. But it's not an explanation. The explanation for God's wanting to be friends with us is presumably much like the explanation for why we want to be friends with some of our fellow human beings. We desire to become friends with someone not because we think

she merits it or because we think her worth requires it, but because we anticipate that our friendship will be a significant good in the lives of both of us. So, too, for God's desire to be friends with human beings.

The Naturalist Response

How might a naturalist respond to my argument; specifically, a naturalist who speaks with the vulgar while thinking with the scientists? Almost certainly, he would first try to falsify my claim that there is no plausible secular grounding of the dignity that accounts for natural human rights; he would try to provide such a grounding. As I have indicated, I think the chance of success in that endeavor is low.

Suppose he fails in that endeavor. What then are his options? One option would be to remove all rights-talk from his use of folk language. Rights-talk is now so deeply embedded in the moral discourse of most human beings that it's hard to imagine doing without it; but, as I indicated at the beginning of this chapter, there have been writers who were not naturalists who have proposed doing exactly that.

Another option for the naturalist would be to continue speaking of natural human rights while conceding that he knows of no grounding for these rights that he finds acceptable. This is not an irrational response. It's a common situation of philosophers and other theorists to firmly believe that such and such is the case without being able to offer an account of the phenomenon. The topic of our discussion is a good example of the point: there are philosophers who firmly believe that there are human rights but find all of the accounts currently on offer unsatisfactory, including the theistic account that I have just now given. An account would be desirable, a good thing to have; affirming the existence of the phenomenon might become shaky at a certain point if no satisfying account was forthcoming. But being unable to account for something that one believes to be the case does not imply that the phenomenon does not exist. On the other hand, sometimes it makes good sense to draw that conclusion.

Instead of responding to my argument in one or another of these ways, could a naturalist instead respond in the way that I hinted at early in the discussion? Could he embrace my theistic account of human rights? Could he introduce the relevant God-talk into his version of folk language? If he can speak of love and worth and rights while believing that love and worth and rights do not really exist, what's to hinder him from speaking of God while believing that God does not exist? I do not know the answer to this question.

References

Ferguson, A. (2013). Who Is Thomas Nagel and Why Are So Many of His Fellow Academics Condemning Him? *Weekly Standard*, Vol. 18, No. 27, March 25.

McLellan, D. (1977). *Karl Marx: Selected Writings*. Oxford: Oxford University Press.

Tierney, B. (1997). *The Idea of Natural Rights: Studies on Natural Rights, Natural Law and Church Law 1150–1625.* Atlanta: Scholars Press.

Wolterstorff, N. (2008). *Justice: Rights and Wrongs.* Princeton: Princeton University Press.

Wolterstorff, N. (2012). *Understanding Liberal Democracy.* Oxford: Oxford University Press.

32

Cognitive and Evolutionary Approaches to Religion

ROBERT N. MCCAULEY

Introduction

David Hume's *The Natural History of Religion* (1976) is probably the best known intellectual ancestor of recent cognitive and evolutionary approaches to religion. These projects share an interest in the origins and forms of religiosity and a broadly naturalistic methodological orientation that inspires caution, if not outright abstemiousness, with regard to related metaphysical and epistemological issues. They also all look to features of the mind as the most important proximate explanatory consideration when accounting for the recurrence of religion among human populations.

Contemporary approaches, however, differ from Hume's work in two respects. They take advantage of significant scientific developments and the explanatory resources that have arisen since Hume's time, and, consequently, they are usually less concerned with the rational status of religious beliefs and actions. Their methods are scientific and their interests are solely explanatory. Therefore, they remain noncommittal with regard to the array of metaphysical assumptions in which religions and their believers routinely traffic, and they accord no privilege to religious knowledge claims or religious "ways of knowing."

Logically, those positions are born of these fields' methodological naturalism (not of some collective antipathy to religion). Certainly, some scholars (Dawkins 2006; Dennett 2006) employ these fields' explanatory proposals and research as part of a larger argument against religion, but more premises than either the cognitive or evolutionary approaches to religion can supply are required for compelling arguments against many religious claims. Nothing in principle or in practice bars people with religious convictions from either contributing to these fields or, of course, making use of these fields' findings as they will. Arguably, evidence from cognitive and evolutionary research about religions generally appears to support the contention that the methodological, epistemological, and metaphysical caution of naturalistic approaches to religion will prove incompatible with the claims of

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

popular religion. Those claims, however, do not exhaust the set of all religious claims, and to presume that that incompatibility extends so far may underestimate the ingenuity of the religious imagination (McCauley 2011).

Setting the Stage

The two most influential scientific developments for these new approaches to religion are the evolutionary and cognitive revolutions. The former is more than 150 years old. In addition to rendering teleological attributions dispensable in the explanation of biological phenomena, Charles Darwin's theory provides a framework for making sense of long-term changes in large-scale (biological) systems, such as populations and species. E.O. Wilson's (1975) sociobiological speculations resuscitated proposals, which Darwin himself entertained (as did William James), that natural selection shapes human behavior no less than human bodies. Wilson's work occasioned in the subsequent decades an array of new evolutionary thinking about human behavior and culture that has variously informed the new scientific theorizing about religion.

The cognitive revolution's origins are more recent and more diffuse. Since the 1950s, works by such figures as Noam Chomsky, George Miller, Ulric Neisser, and Herbert Simon have introduced the possibility of studying the operations of the human mind systematically. That suggestion was revolutionary, because it came after a 30-year reign of behaviorism in experimental psychology. Behaviorism in its several forms maintains that what is available for empirical study is *behavior*, and that mind, if it amounts to anything more than behavior, is inaccessible to empirical scrutiny, let alone experimental investigation. Behaviorism's influence waned in the face of withering criticisms of its limitations (Chomsky 1959), empirical results it could not explain (Garcia et al. 1966), and new findings and explanatory hypotheses spawned by cognitively oriented research (Neisser 1967). The emerging cognitive sciences promised insights about the mental representations, and the mechanisms for their manipulation, that informed thought and action in any sphere.

Different unpromising circumstances across the apposite disciplines (psychology of religion, cultural anthropology, and religious studies) provoked some interested in the explanation of religious beliefs and practices to look to the new cognitive sciences for inspiration. Correlational studies in the psychology of religion pertaining to personality and features of religious experience are certainly worthwhile, but they highlight neither cognition nor action nor causal mechanisms. From Franz Boas to Clifford Geertz, cultural anthropology throughout the 20th century abjured attention to psychology or cognition in deference to the influence of *culture* on conduct and thought, and to its putative autonomy from the exigencies of the workings of individuals' minds. Culture is not the result of those varied workings, but rather the framework shaping them.

Both cultural anthropology and religious studies held that the key to understanding thought and action depended upon fathoming the *meanings* of the cultural and the religious. They concurred that culture and religion are *symbolic*, and, thus, the inquirer's principal task is to ascertain symbols' meanings. Resonant developments within the humanities generally, and in religious studies particularly, involved a decisive turn toward the *primacy of texts*. This position not only holds that texts are the most prominent, plentiful, and lasting evidence about religious thought, especially the religious thought of the past, but that the

methods for interpreting texts' meanings are the principal means for approaching anything that humans find meaningful, including nonlinguistic symbolism.[1]

Dissenting voices submitted that solely searching for meanings of nonlinguistic symbols is inauspicious for explanatory purposes and that cognition deserves more attention than that view allows. Dan Sperber (1975) argued that this general approach, whether in the form of cryptological accounts or of perspectives that appeal to the unconscious, offers no exit from the symbolic.

The cryptological perspective insists that symbols have hidden meanings, in conformity with a code that few, if any, know. Command of such codes allegedly results from recognizing *motivations* informing symbols' connections to meanings. Noting that this standpoint entails that the "mass of humanity obsessively manipulates tools whose usage it does not know, and reiterates messages whose sense it is ignorant of" (1975, 22), Sperber holds that no uniformities underlie the motivations connecting symbols and their hidden meanings, and that putative motivations' apparent plausibility only arises after the fact; that is, *after* learning a pairing between some symbol and its purported meaning. Consequently, assertions about motivations are themselves *symbolic* claims.

Approaches to symbolism appealing to the unconscious, such as Sigmund Freud's, fare no better, says Sperber. Here, the motivation for the connections – in a word, *sexuality*, in Freud's case – is both uniform and clear. The problem is that this approach neither delimits the range of possible symbols (what cannot support some sexual interpretation?) nor articulates the principles that determine the interpretive import of the sexual motivation in specific cases.

Such interpretive approaches supply no tools for gaining explanatory purchase on symbolism, because "all keys to symbolism are part of symbolism itself" (1975, 50). Interpretations are simply *more* symbolism demanding yet more interpretations and so on. Sperber (1975, 112) observes that the symbolic is "not a property either of objects, or of acts, or of utterances, but of conceptual representations that describe or interpret them." Thus, any explanatory leverage with cultural and religious symbolism will come from the study of conceptual representations. Enter cognitive science.

On what they take to be cognitive grounds, E. Thomas Lawson and Robert N. McCauley (1990) grant interpretive materials a somewhat more significant role. Minimally, they regard interpretations as provisionally delineating domains of interest and as employing the categories with which the framing of values and explanatory inquiries must inevitably begin. Their dissent (1993) aims, instead, at assumptions about the primacy of texts and the concomitant primacy of textual tools. Everything looks like a nail to workers whose only tools are hammers. Everything looks like a text to scholars whose only tools are for textual interpretation. But neither rituals nor other religious actions nor temples nor iconography nor flags nor vestments nor crosses nor further nonlinguistic religious symbolism, nor even the conceptual representations that undergird such symbolism, are texts. The literary does not even encompass everything linguistic; indeed, it amounts to but a small fraction of humans' linguistic output. Humans have done a lot more talking than they have reading or writing. Moreover, the interpretation of texts seems misplaced when pondering the religious lives of illiterate people, who have been the vast majority of religious people across history. The emphasis on texts is particularly misplaced when examining religious beliefs and practices in *nonliterate* societies.

1 Their many well-founded moral concerns and their often flamboyant intellectual posturing notwithstanding, the postmodernist heirs to this symbolic-hermeneutic paradigm remain committed to the primacy of texts.

These considerations cut through the smokescreen of abstractions enveloping the doctrine of the primacy of texts in religious studies. The approach is not well suited to analyze religious actions. It is mute about what people must know to take part in coordinated symbolic practices. It is ill equipped to handle the nonliterate origins of religions and the origins and character of religious thought and activity both in preliterate settings and (until the last two centuries) in the minds of the vast majority of religious people.

Given the rhetoric of diversity, inclusiveness, and equality that regularly accompanies work championing textual primacy, especially among postmodernists, it is ironic that Lawson and McCauley (1993) also argue that the doctrine is prejudicial. They contrast the cognitive scientist's interest in the features of the mind that underlie and unite religious thought, expression, and behavior across places and times with the focus of so much work in mainstream religious studies on religious traditions as *rooted in texts*. They contend that this prominence within religious studies of a textually based notion of history and tradition perpetuates colonialist prejudices. Since peoples of nonliterate, small-scale societies have no texts, they cannot be granted traditions or histories in the relevant sense. Their beliefs, activities, and nonlinguistic symbolism remain forever subterranean. Such approaches unavoidably encourage distinctions between more valued (literate, textual) and less valued (nonliterate, nontextual) arrangements, tending, however unintentionally, to favor religions of the book.

The more restrictive conceptions of history and traditions according with textual primacy contrast with Pascal Boyer's (1990) admittedly thinner, cognitively inspired conception of "traditions" as cultural habits and customs. On Boyer's account, participants need not have knowledge about those practices' origins or evolution. Transmission of such behaviors or of representations of their cultural or epistemic prestige suffices for the sense of "tradition" at stake. This notion of tradition (and of history), of course, embraces all of the traditions and histories that the doctrine of textual primacy admits.

Cognitivists ask why nonliterate cultural arrangements should be excluded from scholarly treatments of religions. They disagree that texts and textually based traditions are either all or even the most basic religious phenomena. They maintain that such materials are susceptible to cognitive analyses and that the thinner, cognitive conception of tradition will cast a wider net that captures both textual materials and practices, rituals, spaces, buildings, artifacts, experiences, and so on accorded religious significance. Approaches to religions eschewing the cognitive sciences' theories, findings, and research tools needlessly deprive scholars of resources that:

- Have proven fruitful and effective in other domains.
- Bring greater precision, empirical accountability, and evidential diversity to the study of religions than heretofore.
- Facilitate explicit theorizing and empirical research about a wider range of factors shaping religious thought and conduct.
- Include consideration of pan-human cognitive factors that undergird religious phenomena in *all* cultural settings, regardless of their time, place, size, or technological or intellectual sophistication.

No less than with languages, economies, and social and political arrangements, the cognitive sciences can aid our understanding of religions and their histories. Consequently, archaeologists, theorists of cultural evolution, classicists, scholars of ancient religions (for

which sometimes few, if any, texts are known), and even historians of religion, whose primary materials are texts, are increasingly recognizing the value of the cognitive science of religion (CSR) (Whitehouse and Martin 2004; Pyysiäinen and Uro 2007; Martin and Sørensen 2011; Czachesz and Uro 2013; Richerson and Christiansen 2013).

The Byproduct Theory

The earliest works in CSR (Guthrie 1980; 1993; Lawson and McCauley 1990; Whitehouse 1992; 1995; Boyer 1994) agree that minds do not come automatically equipped with cognitive machinery devoted to religion. Whatever else each position affirms, all declare that religions involve cultural arrangements which effectively engage ordinary cognitive systems that are in place on the basis of considerations having nothing to do with religion or with one another. They also assert that religions are not the only cultural arrangements that do so. Folklore contains comparably amazing personages; militaries can utilize more rituals than religions. Though their ambitions vary and they attend to different cognitive capacities, these early proposals all maintain that these cognitive capacities' exercises in religious contexts are *byproducts* of their normal functioning. Whether they concern anthropomorphism (Guthrie), action representation (Lawson and McCauley), episodic and autobiographical memory (Whitehouse), or all of these and more (Boyer), these cognitive capacities exist in human minds because they enable people to deal with the species' perennial problems.

Stewart Guthrie (1980; 1993) argues that humans' penchant for anthropomorphic representations and explanations of the non-human world is what *all* religions have in common and what sets them apart from other explanatory endeavors. This penchant for anthropomorphism is an all-purpose cognitive tool and is typically humans' *first* resort for managing ambiguous phenomena; that is, things they do not understand. Religions, on his view, primarily deliver intellectual resources for explaining things in terms of the impact on the world of the thoughts and actions of more or less human-like beings. The orderliness, simplicity, and plausibility (as a function of the familiarity) of religions' anthropomorphic accounts fund their allure, which they retain even in the face of overwhelming contrary evidence. Humans find faces in the clouds and not merely meanings but messages in events.

For more than a decade, Justin Barrett (e.g., 2012) has argued that the hypersensitivity of a more general agent detection device (HADD), rather than a penchant for anthropomorphism specifically, is required to explain much that Guthrie's theory aims to encompass. It is not obvious, for example, how a devastating drought *resembles* the ancestors who are presumed to have caused it in order to warn about a community's moral failures. Barrett and Guthrie concur that whatever the precise character of the mechanisms, evolution insured that their settings are acute, given the potentially disastrous consequences of false negatives; better to err on the side of caution.

Lawson and McCauley (1990) advance a more modest speculation, since they only take up participants' knowledge, particularly their *implicit* knowledge, of their religious ritual systems. Developmental psychologists (Rochat, Morgan, and Carpenter 1997) have disclosed that infants as young as 3 months are sensitive to goal-oriented activity and can discriminate between agents and other things, and between actions and other events. Lawson and McCauley assert that people deploy the pertinent cognitive machinery to represent religious ritual actions. Outlining an action representation system that utilizes

no more than elementary assumptions about basic action roles, prerequisite rituals (when they exist), and the assignment of superhuman agents to fulfill one or another of these roles, Lawson and McCauley formulate principles about ritual form that generate systematic predictions about a host of rituals' properties, including their repeatability (with the same participants in the same roles), their reversibility, and their comparative levels of sensory pageantry.

Subsequent research on multiple religious systems (Malley and Barrett 2003) and on the representation of ritual action (Legare and Souza 2012) corroborates the pivotal role that Lawson and McCauley attribute to the positions of the gods as a decisive variable in determining religious rituals' forms. Further research on action representation (Sørensen, Lienard, and Feeny 2006; Nielbo and Sørensen 2011), and on ritualized behaviors generally (Boyer and Lienard 2006), underscores additional cognitive considerations that distinguish rituals and that tie them to humans' intuitions about hazard precautions. Religious rituals are not filled with all of that cleansing, spatial ordering, and meticulous attention to procedures by chance.

Harvey Whitehouse (1992; 1995) explores how contrasting ritual arrangements differ in their mnemonic impact on participants and in their associations with a host of religious, social, and political patterns. The ambition of his overall proposals thus extends far beyond that of the byproduct theory. His two modes of religiosity, the imagistic and doctrinal, are ideal types that delineate two contrasting assemblages of interacting and mutually supportive factors. Different religious arrangements more or less approximate one or the other.

The imagistic mode relies on infrequently performed rituals with comparatively high levels of sensory pageantry and emotional arousal, which both instill salient autobiographical memories and fuse strong psychological ties between the members of small, close-knit groups. The imagistic mode relies primarily on largely unexplicated, loosely linked, nonlinguistic symbolism. Whitehouse has subsequently emphasized the key roles of *dysphoric* ritual and fictive kinship in the imagistic mode (Whitehouse and Lanman 2014).

Whitehouse's doctrinal mode, by contrast, relies on frequently performed, highly routinized rituals with unexceptional levels of sensory pageantry, and on standardized, logically integrated, verbalized formulae. The doctrinal mode emerges with the advent of large-scale societies and the invention of writing (Mullins, Whitehouse, and Atkinson 2013). Here, conviction depends more on intellectual persuasion. The resulting generalized schemas that participants acquire become elements of semantic memory. Doctrinal arrangements befit large-scale societies that must forge a sense of group identification among people who would, otherwise, be strangers.

Quentin Atkinson and Whitehouse (2011) have surveyed ethnographic evidence showing that these two packages of religious, social, and political arrangements correlate, respectively, with different patterns of resource acquisition. The imagistic pattern, for example, arises far less frequently in connection with intensive agriculture than it does with other types of resource procurement.

Boyer (1994; 2001) has provided the most elaborate discussion of the byproduct theory in the course of offering an overall account of human mental architecture (see also Atran 2002; Barrett 2004; Pyysiäinen 2009). Taking inspiration from the evolutionary psychologists (e.g., Tooby and Cosmides 2005), Boyer outlines the provenance and characteristics of the mental systems religions customarily engage. He argues that they are evolved, domain-specific systems for handling problems fundamental to human survival in a natural world in which predation is a given and in a social world that involves competition for resources and mates.

Theory of mind is the mental system that religions most prominently engage, by means of representations of ancestors, gods, angels, saints, demons, and more. Theory of mind is a multifaceted capacity concerned with detecting animate beings and agents, ascertaining their goals and intentions, representing their actions, and reading their minds. Handling the problems of predation requires keen sensitivity to the presence of agents in the environment, to those agents' goals, and to how their goals inform their actions. Managing social competition among humans relies on such abilities as using language, recognizing kin, creating coalitions, protecting reputation, and detecting and coping with things like intrigue, deception, cheating, and gossip. Possessing theory of mind is a decisive advantage in both realms. The ability to ascertain others' goals and intentions enables humans to ponder alternative actions and to anticipate the states of mind and the probable responses of their predators, their prey, and their conspecifics. The same is true when it comes to the gods.

Theory of mind is not the only mental system whose operations religions have evolved to incite. Boyer thinks that particular religions engage this and various other members of an array of evolved, task-specific systems, concerning such matters as language, contamination avoidance, kinship recognition, biological essentialism, and more. Religions develop materials and circumstances that cue such achievements as detecting inspired speech (glossolalia), designating sacred spaces and objects, creating fictive kin, and recognizing inherited authority (e.g., priestly lines).

These systems operate in the mind's basement; that is, they operate mostly below the level of consciousness. Boyer contends that humans are innately prepared to process some kinds of information readily and unreflectively. Those kinds of information correspond to the diverse domains that our species' natural history reveals to be vital to survival and the propagation of our genes. Typically, a few pertinent cues activate these automatic systems. For example, experimental evidence suggests that unconscious recognition of nothing more than eyes in the environment will elicit unconscious inferences about being watched and, thus, produce greater levels of honesty (e.g., Bateson, Nettle, and Roberts 2006). That ready and unreflective information processing includes automatically and effortlessly pursuing any of a vast complement of default inferences appropriate to the disparate materials of the many domains in question – whether inferences about persons' identities on the basis of seeing their faces, or about the underlying syntactic forms of utterances in a natural language (to which someone's linguistic system has been tuned), or about judicious conduct in dealing with a contaminant, or about the character of an animal's potential offspring, and so on.

Such unconscious inferences about "intuitive ontology" (2001) figure centrally in Boyer's explanation of religious representations. Intuitive ontologies constitute foundational theories about the kinds of thing in the world. They are not the products of repeated experience in the relevant domains, as humans spontaneously execute speculative inferences about a huge variety of matters, including causal connections, with regard to which they have little or no direct experience. For example, people instantly draw inferences about animals' innards or their food acquisition patterns on the basis of learning their diets.

However bizarre the assumptions of (other peoples') religions may seem, Boyer asserts that religious ontologies – at least the successful ones – follow a standard pattern. The catalogue of the supernatural is not unconstrained. Boyer proposes that religious concepts violate expectations, usually only one in each instance, associated with some member or other of a small set of intuitive ontological categories, *while preserving* all of that category's further default inferences. The set of intuitive ontological categories consists of ANIMAL, PERSON, TOOL, NATURAL OBJECT, and PLANT. Violations of presumptive

physical, biological, or psychological properties of these categories yield concepts with *counterintuitive* properties, exemplified by walking on water, immortality, and *knowing* other peoples' thoughts, respectively. The violations come in either of two kinds. *Breaches* occur when something transgresses a principle of folk physics, folk biology, or folk psychology that ordinarily applies to it. A person who walks through walls violates a principle of intuitive physics which holds that two objects cannot occupy the same space at the same time. A person born of a wolf breaches our folk biological expectations about the transmission of a species' essential character in reproduction. *Transfers* occur when properties are reassigned to items that do not possess them. Claims about a mountain that is alive transfer a collection of biological properties to a natural object. Claims about a snake that talks transfer a collection of sophisticated psychological capacities to an organism that does not possess them.

Boyer hypothesizes that *minimally counterintuitive* (MCI) concepts enjoy a decisive advantage from the standpoint of selection in the competition among ideas for our attentions. Representations of items (from the list of intuitive ontological categories) that incorporate but a single breach or transfer are MCI.[2] MCI representations are ideas that humans like to think. Boyer submits that they appeal to the human mind because they approximate a cognitive optimum.

First, *all* counterintuitive concepts are *attention-grabbing*. Counterintuitiveness is not the only way to seize attention, but it will suffice. For example, having a conversation with someone is nothing out of the ordinary, but having a conversation with someone who knows exactly what you are thinking (regardless of what you say) is quite another matter.

Second, MCI concepts retain substantial *inferential potential*. An MCI concept's single violation of the assumptions accompanying an intuitive ontological category leaves its abundant inferential power basically intact. Moses may have parted the Red Sea, but we can still infer that he would have made a splash had he jumped in, that his heart was beating throughout the entire episode, and that he expected that the subsequent inundation of the Egyptians would keep his people safe from those pursuers, for the near term at least. These are but three unexceptional inferences which follow from this story that contains the concept PERSON WHO PARTED THE RED SEA, from an innumerable list of inferences that we might draw about physical, biological, and psychological matters, respectively.

Boyer accentuates the instantaneousness and alacrity with which human minds carry out such inferences and the wealth of inferences available. Both contribute to the *relevance* of the representations (Sperber and Wilson 1986), which here is, in effect, a measure of their cognitive appeal. The mind leaps at such representations of agents, particularly when they are embedded in narratives, the same way that a hungry frog leaps at flies. Such ideas are nearly irresistible. Consequently, Boyer maintains:

> many religious phenomena are around because of a conspiracy of relevance…once a particular theme or object triggers rich inferences in a variety of different mental systems, it is more likely to be the object of great cultural attention and elaboration. (2001, 226)

Religious materials produce "snares for thought" (2001, 263).

Gaining someone's attention is necessary for the persistence of an inferentially rich concept, but it is not enough. The *memorability* of MCI concepts is a third consideration

2 For an extended, systematic treatment of this matter, see Barrett (2008).

contributing to their selective advantage. MCI concepts not only fascinate, but tend to stick. People are far more likely to transmit ideas that they remember than ones easily forgotten.

Researchers have explored the empirical merits of claims about MCI concepts' mnemonic advantages. Early studies obtained the predicted effects (Boyer and Ramble 2001; Barrett and Nyhof 2001). In assorted cultural and religious settings on four continents, MCI concepts were remembered significantly better than:

- Normal, intuitive concepts (a person who delivers thoughtful sermons and sleeps at night).
- Substantially counterintuitive concepts that involve more than one violation of intuitive assumptions (a statue that hears and answers prayers, weeps and bleeds, and walks about at night).[3]
- Highly unusual but not counterintuitive concepts (a chocolate table).

Researchers have explored the influence of other variables, such as narrative context (Upal et al. 2007), imagery (Slone et al. 2007), and causal integration (Harmon-Vukić and Slone 2009), but the mnemonic advantage accruing to MCI concepts generally stands, especially with retention intervals measured in months. Recent experiments indicate that the memorability of MCI concepts holds in children as young as seven (Banerjee, Haque, and Spelke 2013).

One response (e.g., Peterson 2013) to these claims about MCI concepts, about religions' engagement of diverse basement-level mental systems, and about the byproduct theory overall, is that they neglect a substantial amount of religious thought and utterance. The byproduct theory may provide insights about the cognitive foundations of aspects of *popular* religious expression and practice, but it has less to say about religious *reflection*, especially as it occurs in literate, large-scale societies, and about its articulation and organization in propositions, sermons, creeds, theologies, doctrines, and so on.

Talking about its basement presumes that the mind has an upstairs too. There, we think consciously, slowly, deliberately, verbally, and often at some remove from immediate circumstances (thus, this cognition is *off-line*). Abstruse theological formulations consistently feature counterintuitive representations, but their counterintuitiveness is copious rather than minimal. Just for starters, the theologically correct Christian God is all-good, all-seeing, all-knowing, all-powerful, and all-present. The theologically informed religiosity of educated participants in large-scale, literate societies regularly employs representations that are *radically* counterintuitive (McCauley 2011).

That this objection arises is no surprise. This is, after all, the kind of religious cognition that is conspicuous in Whitehouse's doctrinal religions, which have by far the largest numbers of adherents, and, of a piece with the doctrine of the primacy of texts, it is the kind of religious cognition on which so much of traditional religious studies has focused. The objection occasions two replies.

First, though, admittedly, not what the byproduct theorists have headlined, the cognitive sciences have also undertaken extensive research on *explicit* cognition. This concern about reflective religious thought discloses a *limitation* of the byproduct view, certainly, but, more

3 An unhelpful convention is referring to these as "maximally counterintuitive" representations (e.g., Norenzayan et al. 2006). Obviously, their number of violations exceeds one, but this hardly makes them *maximally* counterintuitive.

accurately, it identifies further religious phenomena that other resources from the cognitive sciences can address.

Second, empirical findings about the relations between this explicit religious reflection and unconscious cognition, which the byproduct theorists spotlight, suggest that everyday opinion and standard scholarship overestimate the former's importance. Scholars may overplay assumptions about the cognitive preeminence of carefully formulated, theologically correct texts, which commonly issue from years of contestation and debate, for shaping participants' religious understandings.

Barrett and Frank Keil (1996; see also Barrett 1998) furnish evidence that, in on-line tasks, such as processing and recalling narratives, religious people overwhelmingly utilize conceptions driven by the implicit cognition associated with the various unconscious, task-specific systems that appear to underlie so much of popular religion. These authors designed short narratives about interactions between people and God to be consistent with the theologically correct doctrines to which their experimental participants subscribe. Instead of deploying the theologically complex, ecclesiastically approved and policed concepts, which they affirm when questioned about their beliefs, the experimental participants frequently revert to spontaneous, theologically incorrect conceptions in their recollections of these narratives. This is true even when the task is merely to paraphrase, rather than recall, the narratives, when participants have *full access* to the texts. Barrett and Keil obtained such findings with Christians and Jews in America and with Hindus in India.

Jason Slone (2004) outlines evidence of similar patterns in additional religions, and Emma Cohen and Barrett (2008a,b) marshal ethnographic and experimental evidence indicating that theological incorrectness occurs even in unpretentious settings in which scholarly sophistication and ecclesiastical authority are modest. They show that theologically incorrect ideas readily intrude in the thought of followers of a small Brazilian spirit-possession cult. McCauley (2011) argues that theological incorrectness is inevitable. This is an instance of a general pattern in which recurring intuitive assumptions connected with basement-level cognitive systems intrude in thought and can trump painstakingly acquired reflective knowledge, whether theological or scientific.

Religious Cognition as Individually Adaptive

Boyer explicitly ties his view of the mind informing the byproduct theory to the evolutionary psychologists' proposals about the mind's evolution.[4] They state that the mind's numerous unconscious, domain-specific systems that religions engage are adaptations. The byproduct theory is agnostic at best, however, about whether religions' engagements with those systems are ever adaptive. If they are, then as byproducts of those systems' operations, such religious employments would qualify as *exaptations*. Religions coopt mental systems that are adaptations in their own right for different adaptive ends that are wholly independent of the factors that led to those systems' selection originally. Of a piece with functionalists' contentions in the social sciences for the past century that religion appears to further social integration, the leading candidate for an adaptive end that religious cognition might engender is *social cooperation*.

4 McCauley (2011) argues that the byproduct theory requires neither nativism nor muscular versions of modularity.

Among cognitive scientists of religion, Jesse Bering was the earliest advocate for viewing much religious cognition[5] as an adaptation resulting from natural selection. Bering (2006) argues that the scale and complexity of the human *social* environment have constituted selection pressures on the mind's evolution, and he points to a collection of psychological dispositions that children exhibit and that persist in many (probably most) adults that have the collective effect of augmenting those individuals' fitness. These include mental systems that have already been noted, such as the penchants to presume intentional agents in settings in which the evidence is less than compelling, to attribute events to those agents' actions, to construe natural events especially as messages from those agents, to assume that those messages concern moral matters, and to increase honesty and cooperative behavior when being watched.

Bering also discusses beliefs about the mind's persistence after death. He and David Bjorklund provided experimental evidence that both children and adults typically hold views about the mind's continuing desires, emotions, and epistemic states after death and that these beliefs arise independently of acquiring religious beliefs (Bering and Bjorklund 2004). Another psychological trait that Bering highlights is the presence of pervasive teleological assumptions about intentional agents' designs behind nature and individuals' lives (Bering 2011). Deborah Kelemen (1999) has supplied numerous experimental studies supporting the position that children are promiscuously teleological, finding function, purpose, and design throughout the natural world. She has subsequently offered evidence of a similar tendency in adults (Kelemen and Rosset 2009), and even in scientists (Kelemen, Rottman, and Seston 2013). Bering basically concurs with Kelemen that children are "intuitive theists" (Kelemen 2004).

Bering contends that, on average, possessing this collection of (what he takes to be innate) psychological dispositions improves an individual's reproductive success. People who believe that the gods are watching them, that they will have an afterlife, and that the gods will reward or punish them *then* on the basis of their conduct's moral worthiness *now* are people who are more likely to cooperate (Shariff and Norenzayan 2011). Bering speculates that their cooperativeness will tend to boost their reputations and protect them from pernicious gossip. This will improve the probabilities that others will target them for cooperation, thereby allowing them to gain more access to resources and mating opportunities. Bering observes that at least "the illusion of a punitive God assisted their genetic well-being whenever they underestimated the risk of *actual* social detection by other people" (2011, 191). The reproductive success of individuals so equipped cognitively will result in greater frequency of the genes responsible for those dispositions. In large social groups, where the cooperation of strangers is critical to their success, strong social selection forces will favor dependable individuals. The suite of psychological traits that Bering's hypothesis showcases is likely to produce such trustworthy souls.

Though evidence exists for religious participation aiding health and well-being (Koenig, King, and Carson 2012), and, more pertinently, individuals' rates of reproduction (Blume 2009), advocates of religiosity as an adaptation see these patterns primarily as *effects* of the social cooperation religion fosters (e.g., Bulbulia 2006). Joseph Bulbulia and Richard Sosis enlist the tools of signaling theory and niche construction to explain the evolution of

5 Suggesting that much religious cognition may be adaptive need not defeat the byproduct view. Some aspects of religious cognition may be byproducts. Some of those byproducts may be adaptive exaptations. Other features of religious cognition, however, may be outright adaptations arising from natural selection.

humans' penchant for cooperation-enhancing religious displays. Cooperating individuals generally outperform freelancers, but sustaining such cooperation is ordinarily costly. Since undetected cheaters bear no costs, it pays to cheat when detection is quite unlikely, unless cooperation can be supported in a fashion that changes the probabilities of the payoffs. The evolution of cooperation, then, depends upon measures promoting its *stability*.

Signaling theory concerns how "communication, perception, and control systems co-evolve" to manage the associated problems (Bulbulia and Sosis 2011, 364). *Honest signals* enable cooperation by allowing participants to discriminate reliable partners well enough to discourage cheating. Honest signals are hard to fake; that is, they are not easy to produce without the requisite commitment. Expressions of emotions are hard to fake in just this way, and conspicuously advertise on faces information about states of mind. A signal's costliness, such as a large investment of wealth (an engagement ring) or of time (spending days nursing a sick friend), add to its credibility.

Religions traffic in behaviors that seem candidates for honest signals, since they are costly, lack any obvious instrumental aim, and seem to nourish cooperation. Sosis and his colleagues have uncovered evidence that religious displays function as honest signals of individuals' cooperativeness. In examining a large sample of 19th-century American utopian groups, Sosis (2000) finds that the religious communes outlasted secular ones by a significant margin and that the strictest religious groups were the most successful. Sosis and Eric Bressler (2003) show that costly rituals extended the religious communes' longevity but provided no advantage to the secular ones. Using results from a common-pool resource-dilemma economic game as a measure of in-group cooperativeness, Sosis and Bradley Ruffle (2003) supply evidence that the members of Israeli religious kibbutzim on average show significantly greater cooperativeness than members of secular kibbutzim. Contrary to widespread presumptions about the cooperativeness of females and the competitiveness of males, that finding turns exclusively on the religious *males*' performance in the game. Crucially, males in these Modern Orthodox religious kibbutzim participate in *collective* rituals three times each day. (Though the females do comparable numbers of rituals, they do not do so collectively.)

Bulbulia and Sosis (2011) employ *niche construction* to surmount the problems honest signaling faces with large populations of strangers, who need to cooperate, but for whom it is normally impossible to share honest signals directly. Niche construction involves organisms influencing their environments in ways that affect the impinging selection forces, either by altering those environments or by relocating to different ones (Odling-Smee, Laland, and Feldman 2003). Bulbulia and Sosis propose that religions tend to establish *charismatic ecologies*; that is, arrangements that "*compel* relatively powerful and automatic cooperative responses," either negatively through cultivating inattentiveness to the possibilities for breakdowns or positively through learned expectations about the beneficial outcomes. Establishing such perennial arrangements will, for example, create a tendency among proselytizing religions, especially, for standardization of spaces (architecture) and places (pilgrimage sites) and standardization and repetition of symbols and practices across cultural, ethnic, and linguistic boundaries.

Though experimental research (Bering, McLeod, and Shackleforth 2005; Shariff and Norenzayan 2007) indicates that unconscious priming about the gods elicits concerns for reputation and circumspect conduct, such effects seem unable to explain the fact that religious belief is neither necessary nor sufficient for exhibiting cooperation (nonbelievers cooperate in large-scale societies, and under similar circumstances religious people cheat,

on average, about as frequently as everyone else). Bulbulia and Sosis maintain that it is charismatic ecologies, not religious belief *per se*, which sustain cooperation in large groups, as arrangements nurturing cooperation are, in effect, offloaded into the religiously structured social environment. For example, some recent research on synchronous or coordinated activity suggests that it bolsters not only cooperation among participants, but cooperation with people *outside* the group too (Reddish, Bulbulia, and Fischer 2014). On the other hand, armies carry out many synchronous and coordinated activities but are not spontaneously cooperative with out-group members. More importantly, indiscriminate cooperativeness is not obviously adaptive.

Religion and Religious Cognition as Beneficial for Groups

Diverse evolutionary conjectures about religion variously countenance three *mechanisms of selection* – natural, sexual, and cultural – that might operate at any of three different *levels of selection* – genetic, individual, and group (Slone and Van Slyke 2015). Though both focus on natural selection at the individual level, Bulbulia and Sosis explicitly couch their account in terms of *dual inheritance* (2011, 373), which looks to a further criterion for differentiating evolutionary approaches, viz., *units of inheritance*.[6] Dual inheritance countenances genetic and cultural inheritance, and the best known version emphasizes gene–culture coevolution (Richerson and Boyd 2005). Positions that affirm cultural inheritance characteristically appeal to selective forces operating at the group level. These views attend to the consequences of religiously motivated cooperation for religious *groups*, instead of for individuals.

Science, including evolutionary science, regularly fractionates phenomena to gain explanatory leverage on some property, part, or pattern of a larger ensemble (Whitehouse and Lanman 2014). Advocates for selective forces at the group level acknowledge the merits of *all* of the evolutionary hypotheses about religion scouted heretofore (Wilson 2002, 45; Atran and Henrich 2010). They simply aver that selective processes at the group level can contribute to understanding of religiously motivated cooperation especially. Cooperative religious groups, for example, might engulf or eliminate less cooperative groups, or the former might out-reproduce the latter, or the latter might imitate the former.

David Sloan Wilson and Elliott Sober (e.g., 2008) challenge the dominant position in evolutionary thinking, which holds that though theoretically possible, *natural* selection at the level of groups, if it occurs at all, is rare and negligible. Wilson (2002) argues that multilevel selection theory, which includes natural selection at the group level, allows the construal of religious groups as more or less integrated organisms that compete with other groups, with consequences for the fates of group members' genes. Wilson applies the principle that, all else being equal, cooperative groups, as integrated organisms, outcompete less cooperative groups to explain the successes of John Calvin's Geneva during the Reformation, the system of water temples in Bali, the persistence and resilience of Judaism in the face of a history of persecution, and more.

Cultural group selection (CGS), by contrast, stresses *cultural* selection at the group level. Joseph Henrich (2009) maintains that CGS explains recurrent complexes of beliefs and associated credibility-enhancing displays (CREDs), which foster cooperation and

6 Dennett (2006) advocates memes as the units of inheritance, but the theory has generated little progressive experimental research.

commitment to (only some) superhuman agents. According to Henrich, the byproduct theory cannot explain either the commitment or its exclusivity.

Byproduct theorists champion *content* biases in cognition. In elucidating cultural learning, Henrich gives equal heed to *context* biases. Humans get information from other humans, but the information shared by some people, as opposed to others, is clearly more reliable, more likely to be helpful. Henrich thinks that humans possess an evolved bias favoring prestige (Henrich and Gil-White 2001). Children learn the markers of prestige and regard prestigious persons as more worthy of trust and emulation. To decrease possibilities for Machiavellian manipulation, however, cultural learners watch for exemplars who perform CREDs (Henrich 2009, 247). Costly signals are a subset of CREDS, so the empirical evidence for the former transfers. Exemplars eating the foods they recommend are CREDs but not costly signals. According to Henrich (2009, 256), costly signaling theory needs an account of cultural learning and cultural evolution to explain why signaling costliness does not routinely devolve into runaway extravagance.

Henrich constructs a formal model, which indicates that without CREDs, costly practices, including rituals, are not stable and are unlikely to survive, and that vigilance about CREDs can result in stable costly practices. CGS abets rituals that engage humans' contextual biases; those, for example, in which prestigious exemplars publically affirm beliefs (exploiting prestige bias) or undertake practices (scarification, snake handling, etc.) that only committed followers would affirm or undertake (exhibiting CREDs). CREDs are well-wrought to sustain religions' otherwise less plausible counterintuitive beliefs. They help, first, to solve the problem of *commitment* to such beliefs, which, in turn, helps to solidify the group's cooperative arrangements. They also help to explain why participants do not find competing gods tempting.

Ara Norenzayan (2013) has assembled the resources of these many positions and deployed CGS to endorse the pivotal role that the emergence of Big Gods played in solving problems of cooperation in big groups. Big Gods are not tied to a locale. They are so high up that they are, in effect, everywhere all of the time and, crucially, always watching people's morally relevant behaviors and (often) listening to their morally relevant thoughts. Failsafe moral monitors who dispense rewards and, crucially, punishments, Big Gods anchor large-scale cooperation and supply competitive advantages to the religious communities they supervise.

Big Gods expressly link religion and morality – comparatively recently in the species' natural history (de Waal 2013). Norenzayan (2013, 120–121) speculates that, instead of a side effect of agriculture, Big Gods are (partly) responsible for its emergence. On the other hand, he also argues that they can be dispensable. When citizens live in relatively safe, secure environments in societies with adequate material resources that develop governments with trustworthy institutions, legal systems, and police forces that monitor people's conduct in ways that are similar to the gods' oversight, interest in the gods decreases dramatically. Secularized societies indirectly but substantially check people's interests in religion (Talmont-Kaminski 2013).

Trends in Experimental Research

Subsequent research, since the founding of CSR, reflects two further influences in addition to evolutionary theorizing: an interest in experience and brain imaging. These trends together reflect how mainstream the field has become, since they are the three principal influences in cognitive science generally at this time.

The upsurge of interest in religious experience in CSR is ironic, since some practitioners (Boyer 2001) argue that scholars have overestimated its significance. Renewed interest in religious experience is, however, inevitable. That is because many religious people and scholars are convinced that it is special – whether it is or not. Researchers' interest in the character of experience and embodiment is also a general trend across the social, cognitive, and brain sciences. All have developed increasingly effective means for testing hypotheses about the characteristics, causes, and consequences of the varieties of experience, including ones deemed religious (Taves 2009). New methods have spawned noteworthy findings concerning pain thresholds, arousal, empathy, and memory in everyone from Oxford rowers to ritual firewalkers (Cohen et al. 2010; Xygalatas et al. 2011; 2013).

Research with noninvasive brain imaging, especially fMRI, has grown considerably, as the technology has become increasingly available. Brain imaging offers myriad details about the neural structures and activities involved in religious cognition and experience. Patrick McNamara (2009) reviews a diverse collection of neuropsychological and neuroscientific findings bearing on religious experience. He argues that the neural circuitry associated with religious experience tracks that connected with humans' sense of self, suggesting the potential of the former to alter the latter. Uffe Schjoedt and colleagues scanned the brains of Danish Christians while they were hearing prayers recited by people whom they believed possessed religious charisma. The scans showed significant deactivation in their frontal executive networks, which Schjoedt et al. (2011) construed as registering their deference to these purported religious experts. Brain imaging also provides a new body of evidence bearing on longstanding hypotheses about religious cognition. Schjoedt and his colleagues (2009) furnish evidence that their Danish Christian participants' improvisational praying activates brain areas associated with theory of mind and social cognition. For these individuals, such transactions with God are just like any other interpersonal interaction. That, of course, further corroborates the single most fundamental commitment in CSR, viz., that religious cognition is nothing out of the ordinary.

Acknowledgments

I am grateful to Tom Lawson and Kelly James Clark for helpful comments on an earlier draft.

References

Atkinson, Q.D. and Whitehouse, H. (2011). The Cultural Morphospace of Ritual Form: Examining Modes of Religiosity Cross-Culturally. *Evolution and Human Behavior* 32: 50–62.

Atran, S. (2002). *In Gods We Trust*. Oxford: Oxford University Press.

Atran, S. and Henrich, J. (2010). The Evolution of Religion: How Cognitive By-Products, Adaptive Learning Heuristics, Ritual Displays, and Group Competition Generate Deep Commitments to Prosocial Religion. *Biological Theory* 5: 18–30.

Banerjee, K., Haque, O.S., and Spelke, E. (2013). Melting Lizards and Crying Mailboxes: Children's Preferential Recall of Minimally Counterintuitive Concepts. *Cognitive Science* 37: 1251–1289.

Barrett, J.L. (1998). Cognitive Constraints on Hindu Concepts of the Divine. *Journal for the Scientific Study of Religion* 37: 608–619.

Barrett, J.L. (2004). *Why Would Anyone Believe in God?* Walnut Creek: AltaMira.

Barrett, J.L. (2008). Coding and Quantifying Counterintuitiveness in Religious Concepts: Theoretical and Methodological Reflections. *Method and Theory in the Study of Religion* 20: 308–338.

Barrett, J.L. (2012). *Born Believers: The Science of Children's Religious Belief*. New York: Free Press.

Barrett, J.L. and Keil, F.C. (1996). Conceptualizing a Non-natural Entity: Anthropomorphism in God Concepts. *Cognitive Psychology* 31: 219–247.

Barrett, J.L. and Nyhof, M.A. (2001). Spreading Non-natural Concepts: The Role of Intuitive Conceptual Structures in Memory and Transmission of Cultural Materials. *Journal of Cognition and Culture* 1: 69–100.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of Being Watched Enhance Cooperation in a Real-World Setting. *Biology Letters* 2: 412–414.

Bering, J.M. (2006). The Folk Psychology of Souls. *Behavioral and Brain Sciences* 29: 453–498.

Bering, J.M. (2011). *The Belief Instinct: The Psychology of Souls, Destiny, and the Meaning of Life*. New York: Norton.

Bering, J.M. and Bjorklund, D.F. (2004). The Natural Emergence of Reasoning about the Afterlife as a Developmental Regularity. *Developmental Psychology* 40: 217–233.

Bering, J.M., McLeod, K., and Shackelforth, T.K. (2005). Reasoning About Dead Agents Reveals Possible Adaptive Trends. *Human Nature* 16: 360–381.

Blume, M. (2009). The Reproductive Benefits of Religious Affiliation. In *The Biological Evolution of Religious Mind and Behavior*, edited by Eckart Voland and Wulf Schedfenhövel, pp. 117–126. Berlin: Springer-Verlag.

Boyer, P. (1990). *Tradition as Truth and Communication*. Cambridge: Cambridge University Press.

Boyer, P. (1994). *The Naturalness of Religious Ideas*. Berkeley: University of California Press.

Boyer, P. (2001). *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books.

Boyer, P. and Liénard, P. (2006). Precaution Systems and Ritualized Behavior. *Behavioral and Brain Sciences* 29: 595–650.

Boyer, P. and Ramble, C. (2001). Cognitive Templates for Religious Concepts: Cross-cultural Evidence for Recall of Counter-intuitive Representations. *Cognitive Science* 25: 535–564.

Bulbulia, J. (2006). Nature's Medicine: Religiosity as an Adaptation for Health and Cooperation. In *Where Man and God Meet: The New sciences of Religion and Brain*, edited by Patrick MacNamara, pp. 87–121. Westwood: Greenwood Publishers.

Bulbulia, J. and Sosis, R. (2011). Signalling Theory and the Evolution of Religious Cooperation. *Religion* 41: 363–388.

Chomsky, N. (1959). A Review of B.F. Skinner's *Verbal Behavior*. *Language* 35: 26–58.

Cohen, E. and Barrett, J.L. (2008a). Conceptualizing Spirit Possession: Ethnographic and Experimental Evidence. *Ethos* 36: 246–267.

Cohen, E. and Barrett, J.L. (2008b). When Minds Migrate: Conceptualizing Spirit Possession. *Journal of Cognition and Culture* 8: 23–48.

Cohen, E., Ejsmond-Frey, R., Knight, N., and Dunbar, R.I.M. (2010). Rowers' High: Behavioural Synchrony Is Correlated with Elevated Pain Thresholds. *Biology Letters* 6: 106–108.

Czachesz, I. and Uro, R. (eds.) (2013). *Mind, Morality and Magic: Cognitive Science Approaches in Biblical Studies*. Durham: Acumen.

Dawkins, R. (2006). *The God Delusion*. Boston: Houghton Mifflin.

Dennett, D.C. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Viking.

de Waal, F.B.M. (2013). *The Bonobo and the Atheist*. New York: Norton.

Garcia, J., Ervin, F.R., and Koelling, R.A. (1966). Learning with Prolonged Delay of Reinforcement. *Psychonomic Science* 5: 121–122.

Guthrie, S. (1980). A Cognitive Theory of Religion. *Current Anthropology* 21: 181–203.

Guthrie, S. (1993). *Faces in the Clouds*. Oxford: Oxford University Press.

Harmon-Vukić, M.E. and Slone, D.J. (2009). The Effect of Integration on the Recall of Counterintuitive Stories. *Journal of Cognition and Culture* 9: 57–68.

Henrich, J. (2009). The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution. *Evolution and Human Behavior* 30: 244–260.

Henrich, J. and Gil-White, F.J. (2001). The Evolution of Prestige: Freely Conferred Status as a Mechanism for Enhancing the Benefits of Cultural Transmission. *Evolution and Human Behavior* 22: 1–32.

Hume, D. (1976). *The Natural History of Religion and Dialogues Concerning Natural Religion*, edited by Anthony W. Colver and James V. Price. Oxford: Oxford University Press.

Kelemen, D. (1999). The Scope of Teleological Thinking in Pre-School Children. *Cognition* 70: 241–272.

Kelemen, D. (2004). Are Children "Intuitive Theists"?: Reasoning about Purpose and Design in Nature. *Psychological Science* 15: 295–301.

Kelemen, D. and Rosset, E. (2009). The Human Function Compunction: Teleological Explanation in Adults. *Cognition* 111: 138–143.

Kelemen, D., Rottman, J., and Seston, R. (2013). Professional Physical Scientists Display Tenacious Teleological Tendencies: Purpose-Based Reasoning as a Cognitive Default. *Journal of Experimental Psychology: General* 142: 1074–1083.

Koenig, H.G., King, D.E., and Carson, V.B. (2012). *Handbook of Religion and Health*, 2nd edn. New York: Oxford University Press.

Lawson, E.T. and McCauley, R.N. (1990). *Rethinking Religion: Connecting Cognition and Culture*. Cambridge: Cambridge University Press.

Lawson, E.T. and McCauley, R.N. (1993). Crisis of Conscience, Riddle of Identity: Making Space for a Cognitive Approach to Religious Phenomena. *Journal of the American Academy of Religion* 61: 201–223.

Legare, C.H. and Souza, A.L. (2012). Evaluating Ritual Efficacy: Evidence from the Supernatural. *Cognition* 124: 1–15.

McNamara, P. (2009). *The Neuroscience of Religious Experience*. New York: Cambridge University Press.

Malley, B. and Barrett, J.L. (2003). Can Ritual Form Be Predicted From Religious Belief? A Test of the Lawson-McCauley Hypotheses. *Journal of Ritual Studies* 17: 1–14.

Martin, L.H. and Sørensen, J. (eds.) (2011). *Past Minds: Studies in Cognitive Historiography*. London: Equinox.

McCauley, R.N. (2011). *Why Religion Is Natural and Science Is Not*. New York: Oxford University Press.

Mullins, D., Whitehouse, H., and Atkinson, Q.D. (2013). The Role of Writing and Recordkeeping in the Cultural Evolution of Human Cooperation. *Journal of Economic Behavior and Organization* 90: 141–151.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nielbo, K.L. and Sørensen, J. (2011). Spontaneous Processing of Functional and Nonfunctional Action Sequences. *Religion, Brain & Behavior* 1: 18–30.

Norenzayan, A. (2013). *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton: Princeton University Press.

Norenzayan, A., Atran, S., Faulkner, J., and Schaller, M. (2006). Memory and Mystery: The Cultural Selection of Minimally Counterintuitive Narratives. *Cognitive Science* 30: 531–553.

Odling-Smee, J., Laland, K., and Feldman, M. (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton: Princeton University Press.

Peterson, G.R. (2013). McCauley, the Maturationally Natural, and the Current Limits of the Cognitive Science of Religion. *Religion, Brain & Behavior* 3: 141–151.

Pyysiäinen, I. (2009). *Supernatural Agents: Why We Believe in Souls, Gods, and Buddhas*. New York: Oxford University Press.

Pyysiäinen, I. and Uro, R. (eds.) (2007). *Explaining Christian Origins and Early Judaism: Contributions from Cognitive and Social Science.* Leiden: Brill.

Reddish, P., Bulbulia, J., and Fischer, R. (2014). Does Synchrony Promote Generalized Prosociality? *Religion, Brain & Behavior* 4: 3–19.

Richerson, P.J. and Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Richerson, P.J. and Christiansen, M.H. (eds.) (2013). *Cultural Evolution. Strüngmann Forum Reports*, vol. 12. Cambridge, MA: MIT Press.

Rochat, P., Morgan, R., and Carpenter, M. (1997). Young Infants' Sensitivity to Movement Information Specifying Social Causality. *Cognitive Development* 12: 441–465.

Schjoedt, U., Stødkilde-Jørgensen, H., Geertz, A.W., and Roepstorff, A. (2009). Highly Religious Participants Recruit Areas of Social Cognition in Personal Prayer. *Social Cognitive and Affective Neuroscience* 4: 199–207.

Schjoedt, U., Stødkilde-Jørgensen, H., Geertz, A.W., Lund, T.E., and Roepstorff, A. (2011). The Power of Charisma – Perceived Charisma Inhibits the Frontal Executive Network of Believers in Intercessory Prayer. *Social, Cognitive, and Affective Neuroscience (SCAN)* 6: 119–127.

Shariff, A.F. and Norenzayan, A. (2007). God Is Watching You: Priming God Concepts Increases Prosocial Behavior in an Anonymous Economic Game. *Psychological Science* 18: 803–809.

Shariff, A.F. and Norenzayan, A. (2011). Mean Gods Make Good People: Different Views of God Predict Cheating Behavior. *International Journal for the Psychology of Religion* 25: 85–96.

Slone, D.J. (2004). *Theological Incorrectness: Why Religious People Believe What They Shouldn't*. New York: Oxford University Press.

Slone, D.J. and Van Slyke, J.A. (2015). *The Attraction of Religion: Evolutionary Theories of Religion*. London: Bloomsbury.

Slone, D.J., Gonce, L.O., Upal, A., Tweney, R., and Edwards, K. (2007). Imagery Effects on Recall of Minimally Counterintuitive Concepts. *Journal of Cognition and Culture* 7: 355–367.

Sørensen, J., Liénard, P., and Feeny, C. (2006). Agent and Instrument in Judgments of Ritual Efficacy. *Journal of Cognition and Culture* 6: 463–482.

Sosis, R. (2000). Religion and Intragroup Cooperation: Preliminary Results of a Comparative Analysis of Utopian Communities. *Cross-Cultural Research* 34: 77–88.

Sosis, R. and Bressler, E. (2003). Co-Operation and Commune Longevity: A Test of the Costly Signaling Theory of Religion. *Cross-Cultural Research* 37: 11–39.

Sosis, R. and Ruffle, B.J. (2003). Religious Ritual and Cooperation: Testing for a Relationship on Israeli Religious and Secular Kibbutzim. *Current Anthropology* 44: 713–722.

Sperber, D. (1975). *Rethinking Symbolism*, translated by Alice Morton. Cambridge: Cambridge University Press.

Sperber, D. and Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Talmont-Kaminski, K. (2013). *Religion as Magical Ideology: How the Supernatural Reflects Rationality*. Durham: Acumen.

Taves, A. (2009). *Religious Experience Reconsidered: A Building-Block Approach to the Study of Religion and Other Special Things*. Princeton: Princeton University Press.

Tooby, J. and Cosmides, L. (2005). The Conceptual Foundations of Evolutionary Psychology. In *Handbook of Evolutionary Psychology,* edited by David Buss, pp. 5–67. New York: John Wiley & Sons.

Upal, A., Gonce, L.O., Slone, D.J., and Tweney, R. (2007). Contextualizing Counterintuitiveness: How Context Affects Comprehension and Memorability of Counterintuitive Concepts. *Cognitive Science* 31: 415–439.

Whitehouse, H. (1992). Memorable Religions: Transmission, Codification and Change in Divergent Melanesian Contexts. *Man* 27(4): 777–797.

Whitehouse, H. (1995). *Inside the Cult: Religious Innovation and Transmission in Papua New Guinea*. Oxford: Clarendon Press.

Whitehouse, H. and Lanman, J. (2014). The Ties that Bind Us: Ritual, Fusion, and Identification. *Current Anthropology* 55(6): 12.2014.

Whitehouse, H. and Martin, L. (eds.) (2004). *Theorizing Religions Past: Archaeology, History, and Cognition*. Walnut Creek: AltaMira.

Wilson, D.S. (2002). *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*. Chicago: University of Chicago Press.

Wilson, D.S. and Sober, E. (2008). Evolution "for the Good of the Group." *American Scientist* 96: 380–389.

Wilson, E.O. (1975). *Sociobiology*. Cambridge, MA: Harvard University Press.

Xygalatas, D., Konvalinka, I., Roepstorff, A., and Bulbulia, J. (2011). Quantifying Collective Effervescence: Heart-Rate Dynamics at a Fire Walking Ritual. *Communicative and Integrative Biology* 4: 735–738.

Xygalatas, D., Schjoedt, U., Bulbulia, J., Konvalinka, I., Jegindø, E.-M., Reddish, P., Geertz, A.W., and Roepstorff, A. (2013). Autobiographical Memory in a Fire-Walking Ritual. *Journal of Cognition and Culture* 13: 1–16.

33

The Naturalness of Religious Belief: Epistemological Implications

HELEN DE CRUZ

> No passions, therefore, can be supposed to work upon such barbarians, but the ordinary affections of human life; the anxious concern for happiness, the dread of future misery, the terror of death, the thirst of revenge, the appetite for food and other necessaries. Agitated by hopes and fears of this nature, especially the latter, men scrutinize, with a trembling curiosity, the course of future causes, and examine the various and contrary events of human life. And in this disordered scene, with eyes still more disordered and astonished, they see the first obscure traces of divinity.
>
> Hume (1757, 14–15)

Naturalism, Cognitive Science of Religion, and the Reasonableness of Religious Beliefs

In recent years, analytic philosophers have been increasingly drawn toward naturalism. They attempt to bring philosophical claims, for instance about knowledge, the human mind, and ethics, in line with findings from the sciences. Naturalistically inclined philosophers want to explain or reduce phenomena to natural (as opposed to supernatural) causes. Alvin Goldman (1999) has developed what he terms a "moderate naturalistic" position, where epistemic warrant or justification[1] is a function of the psychological processes that produce or preserve belief. To understand such processes, Goldman recommends that philosophers rely on findings from the cognitive sciences, such as neuroscience and cognitive psychology. In other words, if we want to know whether particular beliefs we hold are

1 There is a continued debate in epistemology on what makes beliefs warranted or justified, and what these terms mean. I will not attempt to define them here, but will use them in a loose sense to denote states that give beliefs a positive epistemic standing.

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

warranted, we need to get a better understanding of the psychological processes that underlie these beliefs, and cognitive science can assist us in this.

The question of whether religious beliefs are warranted or justified has been an enduring topic of debate in philosophy. Can people ever be justified in believing in God? Over the past few decades, a new scientific approach to religion has arisen. The *cognitive science of religion* (CSR) is the multidisciplinary study of religious beliefs and practices, incorporating findings from anthropology, evolutionary theory, and developmental and cognitive psychology, among other fields. Like other cognitive scientists, CSR scholars aim to understand the human mind by examining how we process information through the study of cognitive processes like perception, memory, and reasoning. If, as Goldman argues, the psychological processes that underlie beliefs are indeed relevant for their epistemic status, naturalistically inclined philosophers of religion should take findings from CSR seriously.

In this chapter, I will examine the implications of CSR for the epistemic standing of religious beliefs by focusing on the naturalness of religious beliefs. Many cognitive scientists of religion, such as Paul Bloom (2007) and Robert McCauley (2011), have characterized the belief processes and practices that underlie religions across the world as "natural"; I will refer to this as the naturalness of religion thesis (NRT). I start out by tracing the origins of NRT in early modern natural histories of religion. These works traditionally had an antitheistic agenda, which partly explains why CSR is still regarded as a project that has mainly negative implications for the rationality of religious beliefs. Next, I will briefly discuss assumptions underlying CSR. I will then analyze the term "natural" as it is used in CSR, focusing on McCauley's concept of maturational naturalness. In the final section, I will argue that, because of CSR's self-imposed methodological limitations, NRT as it is currently developed in CSR does not have negative implications for the epistemic standing of religious beliefs.

Natural Histories of Religion

CSR, a relatively recent research program that originated in the 1980s, has ancient roots. Since Antiquity, philosophers have attempted to seek the natural origins for belief in supernatural beings. For example, Xenophanes thought that religious imagination is constrained by how we conceptualize other human beings: "If cattle and horses and lions had hands, or could paint with their hands and make art like humans can, horses would paint or sculpt gods resembling horses and cattle would make them resemble cattle, each according to how they look" (6th century BCE/1903, 54, author's translation). Euhemerus (late 4th century BCE) believed that the gods were deified renowned ancestors and rulers, and that myths have their roots in historical events that have, through retelling, become increasingly exaggerated.

The most direct intellectual precursors to CSR are 17th- and 18th-century natural histories of religion.[2] Natural history differs from contemporary scientific practice primarily in its methodology: rather than testing hypotheses through controlled experiments, natural historians formulated a causal story that they supported by reports and anecdotes from various sources, which were collected in a loose and unsystematic way (e.g., Buffon 1766).

2 The term "natural history" does not correspond to any scientific discipline in its current form, though it is most closely associated with the life and earth sciences, as is still reflected in the term "natural history museum."

Next to this descriptive component, natural historians also had a metaphysical agenda: they wanted to provide naturalistic explanations for domains of human behavior and culture. This agenda is clearly discernible in authors such as d'Holbach and de La Mettrie, who sought to explain domains like religion, emotions, and morality in terms of physical, nonsupernatural causes (Wolfe 2009).

Sociologist of religion Rodney Stark (1999) has argued that natural histories of religion were implicitly or explicitly motivated by an antitheistic agenda. Authors like Bernard Le Bovier de Fontenelle were not so much concerned with writing objective accounts of the origin of religion as with tracing its roots to irrational beliefs. While these natural histories ostensibly aimed at discrediting indigenous religions (which the authors only knew about through hearsay) and pagan religions, Christianity often formed the real target.

For example, de Fontenelle's *Histoire des oracles* (1728) recounts the historical origins of pagan oracles, and indirectly criticizes all religious beliefs – just as the consulters of oracles were irrational in believing their forecasts, so too are contemporary Christians for believing in miracles and other supernatural phenomena. In *De l'origine des fables* (1724/1824), he contends that the earliest humans, ignorant of causes in their environment, invented supernatural phenomena to account for them, as people are naturally attracted to stories with wonderful and miraculous elements. If one has no explanation for a given phenomenon, the supernatural becomes the default explanation: "To the extent that one is more ignorant, or one has less experience, one sees more miracles"[3] (1724/1824, 295, author's translation). Myths are thus created to provide ad hoc explanations for otherwise inexplicable phenomena. The history of myths could be summed up as "a history of the errors by the human mind"[4] (1724/1824, 310, author's translation). According to de Fontenelle, the falsehoods of myths gradually became replaced by factual accounts of past events thanks to the slow development of reason in Western civilization. The same would happen in other cultures, given enough time.

Hume's *Natural History of Religion* (1757) is a late offering in the natural history tradition. Probably owing to his reputation in philosophy, this work is often read in isolation, but it fits very well in the French tradition of natural histories of religion (see also De Cruz in press for a contextual discussion). Like his predecessors, Hume thought that religious beliefs are rooted in features of human psychology, in a way that makes the rationality of these beliefs doubtful. He combined, in a piecemeal and haphazard fashion, his knowledge of Greco-Roman, Egyptian, and other historical religions with second-hand reports of non-Western religions to construct his natural history, which can be summarized as follows. Historically, humans were ignorant about natural causes of occurrences like famines, droughts, and diseases. They feared such events, and were strongly motivated to gain some degree of control over their environment. As a result, they anthropomorphized their surroundings by inventing spirits, gods, and other supernatural beings that could be placated, bribed, or cajoled into doing what humans desired. In line with their own political organizations, they tended to endow one of these spirits or gods with more power than the others, leading to some form of henotheism. Eventually, in several traditions, this important deity overshadowed all the others, leading to monotheism. However, the cognitive inclination toward polytheism remains very strong, as can be seen, for instance, in the reverence of saints in Roman Catholicism.

3 *A mesure que l'on est plus ignorant, et que l'on a moins d'expérience, on voit plus de prodiges.*

4 *Ne cherchons donc autre choses dans les fables, que l'histoire des erreurs de l'esprit humain.*

Did Hume mean to criticize *all* forms of religious belief, or just what he called the "gross polytheism and idolatry of the vulgar" (1757, 35)? Some Hume scholars, like Peter Kail (2007), argue for the former. By contrast, Lee Hardy (2012) writes that Hume's natural history outlines two roads to monotheism: the high road of philosophical reflection and the observation of design in nature, and the low road of haphazard, irrational thought processes. Benjamin Cordry (2011) interprets Hume as an atheist who was nonetheless open to the intellectual virtues of a thin, philosophically informed theism. However, even if Hume was open to rational theism, he thought that most people, even those who believe in the God of Christianity, are unjustified in their beliefs: "Even at this day, and in *Europe*, ask any of the vulgar, why he believes in an Omnipotent Creator of the world; he will never mention the beauty of final causes, of which he is wholly ignorant" (1757, 42). To Hume, the conclusion that the religious beliefs of the vast majority of people are unjustified is inescapable:

> in all nations, which have embraced polytheism or idolatry, the first ideas of religion arose not from a contemplation of the works of nature, but from a concern with regard to the events of life, and from the incessant hopes and fears, which actuate the human mind. (1757, 14–15)

During the 19th century, the sciences became increasingly specialized. Disciplines such as anthropology, sociology, and psychology of religion replaced natural histories of religion. Yet, even in these fields, the scientific study of religion was rarely motivated by a dispassionate search for the psychological and social processes involved. The assumption that religious belief had its roots in non-truth-tracking psychological processes continued to drive theorizing. Auguste Comte (1841), for example, saw the intellectual progress of humanity as proceeding through three stages. The theological state (*état théologique*) is the most primitive of these, starting out with fetishism, and gradually evolving from polytheism to monotheism. The metaphysical stage depersonifies God, and conceptualizes him as an abstract being. Eventually, this abstract God concept gives way to the positive stage, where scientific explanation has replaced religion, and where scientific inquiry is the sole means of regarding the world. Émile Durkheim (1915) saw religion as a set of imaginary beliefs that serve as a social glue; the only factuality of religion is its social nature. Sigmund Freud (1927) considered religion as an illusion, a childlike longing for a father figure caused by a feeling of helplessness. His account was one of the last of its kind, a grand explanation for religion which was not grounded in scientific observations, and was explicitly formulated to discredit religious beliefs.

In the course of the 20th century, the scientific study of religion became more descriptive. Anthropologists like Bronislaw Malinowski (1925/1992) and Edward Evans-Pritchard (1937/1965) began to study the religious beliefs of non-Western people from small-scale societies *in situ*, rather than relying on second-hand reports. They found that religious beliefs in these cultures had a high degree of sophistication. For instance, Malinowski noticed that Trobriand Islanders invoked magic only for deep-sea fishing, a mode of subsistence that was risky and had uneven returns, but not for everyday lagoon fishing, which was safe and had stable returns. Evans-Pritchard found that the Azande (Central Africa) were well aware of natural causes, such as termites, that might cause a house to collapse. They invoked witchcraft to explain why a particular house has collapsed, and reasoned that witches might well use termites and other natural causes to accomplish their ends.

It became glaringly obvious that earlier studies of religion were not only based on incorrect, highly biased ethnographic accounts, but were also racist. For instance, Durkheim

(1915) paid special attention to Australian aboriginal religion (in particular, totemism) because he believed Aboriginals were the most primitive people, the lowest point on a linear scale of civilization and mental development. However, fieldwork demonstrated both the particularity and the complexity of non-Western small-scale societies – for example, Aboriginals have their own complex religious beliefs, such as the Dreamtime, which have no Western parallel. With the disappearance of the myth of the primitive mind, and an increasing recognition of religious diversity and complexity across cultures, anthropologists and other social scientists of religion abandoned their attempts to provide a unified history of religion through successive stages that all cultures have to progress through.

Cognitive Science of Religion

Renewed attempts to explain religion in its entirety started in the 1980s, with the anthropologist Stewart Guthrie's programmatic essay "A Cognitive Theory of Religion" (1980) and his later monograph *Faces in the Cloud*s (1993). Guthrie proposed that religious beliefs are products of anthropomorphism, a common and universal disposition in human cognition. Humans have an overactive tendency to see anthropomorphic faces and agency in their environment, a "reasonable illusion with highly general causes" (1980, 182) that has adaptive benefits: it's better to mistakenly "observe" an agent than to fail to discern one that is present. This idea has been further explored by Justin Barrett (2004), who proposes that religious beliefs are outputs of a hyperactive agency detection device (HADD), a hair-trigger propensity to see agents everywhere in the environment.

CSR authors, in books like *Religion Explained* (Boyer 2002) and *Why Would Anyone Believe in God?* (Barrett 2004), have sought to explain religious beliefs by appealing to universal and normally developing features of human cognition. In this respect, CSR resembles natural histories of religion in its aspiration to provide a comprehensive account of religion. However, unlike natural histories of religion, work in CSR is supported by controlled empirical observations and detailed, first-hand anthropological studies. Moreover, contrary to natural histories of religion and their immediate successors, the identification of causes for religious belief in CSR is descriptive rather than normative. While natural historians of religion hoped to contribute to a more rational frame of mind by exposing the unreliable, primitive origins of all religious beliefs, CSR identifies mundane cognitive processes shared by humans across cultures.

One of the key assumptions of CSR is that religious beliefs and practices are not purely the result of cultural dynamics but are partly shaped by universal and stable features of human cognition, which can be discovered by studying thought processes in young children. They assume that young children are less influenced by cultural expectations and schemas than adults; their spontaneous beliefs reveal central and early-developed features of human cognition. For example, in a classic experiment that probed the origins of afterlife beliefs, preschoolers and elementary school children were shown a puppet alligator eating a toy mouse (Bering and Bjorklund 2004). Though the children understood that the mouse was dead, and could not eat, drink, or breathe anymore, they believed it still had emotions and desires (e.g., the dead mouse still loves its mother). Interestingly, preschoolers made more continuity responses than older children, indicating that children do not intuitively equate the self with the body, but with something nonmaterial. On the basis of this and other experiments, Bloom (2007) concludes that children are intuitive substance dualists.

If children spontaneously believe that mental states can survive the physical death of an agent's body, it is unsurprising that so many cultures have afterlife beliefs, such as the belief that the soul can reincarnate in a different body.

Most CSR authors assume that the human mind is not an undifferentiated blank slate; rather, it is composed of several specialized inference systems, which are termed "intuitive ontologies" (see De Cruz and De Smedt 2015 for an overview). These systems help us gain knowledge about different aspects of the world, such as how inanimate objects move (intuitive physics), how plants and animals grow and behave (intuitive biology), and the mental states of others (intuitive psychology). On this standard picture, we are naturally disposed to think about minds and bodies as separate entities, and religious beliefs are a byproduct of the way our brains have evolved. For example, intuitive dualism emerges as a result of an inherent tension between different inference systems for physical objects and mental states. The inferences of our intuitive psychology do not require or stipulate a physical object (the human body) to perform the cognizing, making it difficult to grasp that a person's mental states have stopped once his or her physical body no longer functions. Though humans have physical bodies, our intuitive psychology makes it quite easy for us to imagine what a deceased agent would think about a state of affairs, as is clear in phrases like "Grandma wouldn't have liked this," or "It's what she would have wanted." CSR is at present a highly successful and proliferating research program. Examples of topics of investigation include rituals, intuitive creationism, the relationship between folk religious beliefs and theology, prayer, belief in supernatural beings, souls and other nonphysical parts of a person, and belief in the afterlife.

The Concept of Naturalness in CSR

CSR authors frequently claim that religion is a natural phenomenon, as is illustrated by titles such as *Religion is Natural* (Bloom 2007) and *Why Religion is Natural and Science is Not* (McCauley 2011). In these studies, the term "natural" is used in at least three distinguishable senses: metaphysical naturalness, phenomenological naturalness, and developmental naturalness. Together, these three senses constitute the NRT.

Metaphysical naturalness is the claim that religion is the result of natural, as opposed to supernatural, processes. This view continues in the tradition of 17th- and 18th-century natural historians, who reduced religion to nonreligious psychological dispositions and social factors. For example, Guthrie (1993) explains belief in supernatural beings by appealing to a cognitive mechanism which has as its evolved function the detection of natural beings, viz., humans and other agents. Similarly, Deborah Kelemen's (2004) hypothesis that children are intuitive theists hinges on the claim that children are naturally prone to see purposiveness and design in nature. This "promiscuous teleology," which gives rise to belief in supernatural creators, has a natural origin: our ability to successfully understand the designer intentions of artifacts. Bloom's (2007) intuitive dualism, likewise, does not originate from an interaction with nonphysical minds, but from an intuitive distinction between minds and bodies.

Phenomenological naturalness refers to the intuitive plausibility and ease with which religious beliefs are acquired and held. For example, Barrett (2010) argues that religion is a mode of thinking that emerges spontaneously as a result of the workings of our cognitive capacities. Folk religious ideas are easy to come up with and to grasp, and require little cognitive effort.

This claim seems plausible when we look at the global distribution of religious beliefs: 85–90% of the world's population believes in one or more gods (Zuckerman 2007). By contrast, atheism is rare and requires cultural support, such as a well-developed secular worldview.

Developmental naturalness is the claim that religious beliefs, such as belief in God, souls, and an afterlife, emerge early, as a spontaneous output of normal cognitive functions that become operational early in development. McCauley (2011) emphasizes this aspect of the NRT. He distinguishes between two kinds of naturalness: maturational and practiced. This distinction teases apart two developmental trajectories of skills and practices that we perform effortlessly, such as cycling, running, and surmising what other people desire or believe. While these skills are phenomenologically natural, they come about through two radically different developmental trajectories. Given that McCauley's account is one of the most detailed conceptual analyses of the naturalness of religion to date, I will critically discuss it in detail here.

Speaking one's mother tongue, walking, and chewing are what McCauley terms *maturationally natural skills*. They are culturally ubiquitous and arise early in human development without explicit instruction. They do not require institutions or material support – as a matter of fact, infant walkers do not speed up walking, and educational DVDs for babies do not speed up infant language development. Such skills are acquired through everyday interactions. Children acquire them with ease, and spontaneously engage in them. While there is natural variability in maturationally natural skills (e.g., some people are better in speaking their first language than others), there is no clear notion of expertise. Maturationally natural skills typically address central human challenges and concerns, such as navigating one's environment or communicating with others.

By contrast, *practiced natural skills*, like reading, driving, and playing a musical instrument are more restricted across times and cultures. They often require years of effortful training, supported by teachers, and material culture. They do not arise as a result of species-typical cognitive development. In fact, they often subtly alter the brain's anatomical structure. For example, the brains of people who have learned to read as adults show structural differences in connectivity compared to those of adults from the same community who learned to read as children (Carreiras et al. 2009). Because of their reliance on sustained practice, practiced natural skills have experts: people who are recognized by others in the field as proficient, such as chess grandmasters, professional musicians, and calligraphers.

McCauley contrasts ordinary religious beliefs and practices with theology and science: the former are maturationally natural, whereas the latter are practiced natural only to the select few (theologians and scientists) and completely unnatural to the majority of human beings. This explains why religion is cross-culturally ubiquitous, and easily acquired, whereas science and theology are culturally restricted, requiring extensive material and institutional support. McCauley offers glossolalia (speaking in tongues) as an example of a maturationally natural skill. This is a religious practice that draws extensively upon natural language, a maturationally natural skill set. While the utterances may sound like a foreign language, upon closer scrutiny they appear to draw exclusively on phonemes and stress patterns that occur in the language of the utterer. The speech-like patterns spontaneously elicit a quest for meaning and interpretation in listeners, precisely because they draw on our maturationally natural capacities to interpret language.

McCauley's claim that religion is maturationally natural should be nuanced. Glossolalia is not cross-culturally universal, but appears in only a few religious traditions, such as

Pentecostal and Charismatic Christianity. It requires specific techniques and a form of expertise; according to its practitioners, speaking in tongues and knowing how to interpret such speech are gifts of the Spirit, which are not manifest in everyone. The Internet offers several detailed tutorials on how to speak in tongues – if glossolalia are so natural, why do people need guidance on how to produce them? Presumably, being part of a culture where speaking in tongues is a common practice provides the exposure necessary to speak in tongues. In non-Pentecostal churches, parishioners do not spontaneously speak in tongues. For example, the Episcopalian minister Lauren Winner (2006, 257–258) recalls in her autobiography how for several months she tried it in vain.

A religious practice like glossolalia seems more akin to music (a practiced natural capacity) than to language. Though musical scales could divide an octave into infinitely many different parts, cross-culturally most of them fall within a small set of scales, comprising only five to seven tones. Scales that are most in line with harmonic series (which correspond to human vocal fold vibrations) are most prevalent (Gill and Purves 2009). Understanding melodies thus depends on maturationally natural capacities, involved in human voice discrimination. Nevertheless, understanding musical structure and understanding harmony are practiced natural skills. In a series of experiments, David Huron (2006) asked Balinese and American musicians to guess the next note in a piece of gamelan music. The best American musicians were less accurate than the weakest Balinese – though they performed far above chance, learning to predict gamelan tunes rapidly. Presumably, the American subjects did better than chance and were such quick learners because gamelan music depends on a constrained set of musical scales. These align closely with harmonic series, to which our ears are naturally attuned. Music critically depends on maturationally natural capacities, but extends them radically as well; it is both practiced and maturationally natural (see De Cruz and De Smedt 2015 for a further development of this point).

The same can be said about other religious beliefs and practices. Even if the cognitive underpinnings of religious beliefs are maturationally natural, children do not spontaneously generate fully-fledged religious systems like Zoroastrism or Christianity. Anthropologists read up extensively on the religious systems they will examine in the field, and they immerse themselves in the culture for weeks or, preferably, months to come to grips with the religions they study. In her recent fieldwork, anthropologist Tanya Luhrmann (2012) provides a detailed case study of the practiced naturalness of ordinary religious belief in the Vineyard Church, communities of Evangelical Christians in Chicago and Palo Alto. The aim of her work is to examine how people learn to internalize and experience as real the religious beliefs of the community they are part of. She demonstrates that the religious experiences of Evangelical Christians require a radical and counterintuitive reconceptualization of one's theory of mind: they have to accept their minds as porous, with some of their thoughts not their own, but coming from God. It is difficult to discern, however, which thoughts are caused by God and which aren't. This requires extensive practice, such as imagining God as a friend who is sitting next to you during prayer. These communities extensively rely on material culture, such as a special room (or, in one instance, a special closet) used for prayer, an object to fixate on while praying (e.g., a candle), and prayer journals (for tracking fleeting thoughts). In this material culture, the Bible occupies a central place: it serves as a springboard of communication with God, as some passages jump out and speak to the reader. In the Vineyard Church, some people are recognized as experts in prayer: they can discern God's voice with more clarity than others, and occasionally hear clearly spoken messages. Such experts are typically congregation members with many years of prayer

experience. The claim that prayer is hard work is not only put forward by Luhrmann's interlocutors. For instance, the 16th-century mystic Teresa of Ávila describes in *The Interior Castle* a combination of contemplative prayer, meditation, and other mindful practices that prepare the mind to become receptive to God.

Religious belief, at least as experienced by Christians, does not seem to fit the description of maturational naturalness. With its reliance on material culture, the importance of continuous and sustained practice, and its recognition of experts, it is better described as a practiced natural skill, with roots in maturationally natural capacities (presumably, most practiced natural skills build on maturationally natural capacities). As CSR indicates, conceptualizing a supernatural being (God) who is concerned with our moral conduct and who has special knowledge and powers is maturationally natural. Learning to effectively communicate with a triune God is a practiced natural skill.

NRT and the Rationality of Religious Belief

How does the NRT impact the rationality of religious beliefs? To examine this question, I will look at the three senses of naturalness discussed earlier, and how they relate to the epistemic standing of religious beliefs. Metaphysical naturalness has frequently been marshaled as a way to debunk religion. Once we regard religion as a natural phenomenon, Dennett (2006) contends, we can lay the idea that religion lies beyond the scope of scientific investigation to rest. "Breaking the spell" means finally opening religion up to scientific scrutiny; something that has been hampered by religious people.[5] Many scholars reduce religion to nonreligious causes, denying that religiosity itself has any effect at all. For example, the observation that religious people are happier or healthier is explained by the fact that belonging to a church provides one with a social network of friends one can rely on and socialize with. Religiosity is thus reduced to nonreligious variables, such as belonging to a social network (Stark 2000).

The scientific study of religion presents a picture in which the religious believer can scarcely recognize herself. Brad Gregory recounts his attempts to understand why Mormons (adherents to the Church of Jesus Christ of the Latter-Day Saints, LDS) believe and live as they do:

> However satisfactory my explanations were to me, or to others who shared my assumptions, those explanations failed to help me understand the members of the LDS Church any better… In fact, because such explanations are inherently reductionist, in the literal sense of reducing religion to something else, they led *away* from any sense of what Mormon experience, beliefs, and behaviors were like. As the colloquial phrase goes, I did not "get" Mormons, not despite but *because* of my explanations, since all of them implied that Mormonism is not what Mormons take it to be. (2006, 133–134)

This reductionist paradox – explaining religion by reducing it to nonreligious causes, and thereby removing what has to be explained – also affects CSR. It precludes understanding religion *as* religion. Take, for instance, Bloom's explanation of belief in the afterlife. Does his

5 Curiously, this claim disregards the fact that religion has been the subject of scientific study since at least the 17th century. In fact, we see that metaphysical naturalism has always been a prominent feature of natural histories of religion.

intuitive mind–body dualism say anything meaningful about the hope for the general resurrection among Christians; that is, the belief that human beings will rise from the dead and have a renewed physical body? Or about the belief that reincarnation is an endless cycle of suffering, from which one would want to break free? CSR provides an explanation only for dispositions to believe in gods, ghosts, and souls, but does not explain why, in particular cultures, people believe in beings like the God of Abraham, Shiva, or the Rainbow Serpent. Thus, CSR "explains" only a very small subset of religious beliefs. For example, Boyer (2002) is a study of memory and transmission biases: it examines which beliefs are most memorable, which in turn explains how religious beliefs get spread, but it says relatively little about their origins.

Metaphysical naturalists using CSR data might respond that though CSR explanations are not exhaustive in detail, and do not adequately account for thick, elaborate religious beliefs, they nevertheless provide convincing naturalistic accounts. Appealing to parsimony, they could argue that there is no need for a supernatural explanation. However, given that CSR is methodologically naturalistic, it is unsurprising that it uncovers only natural explanations for religion. Imagine a scientific discipline that studies body mass in animals and which excludes *a priori* the possibility that body mass could be influenced by sexual selection. Unsurprisingly, its practitioners will find that sexual selection plays no role whatsoever in determining body mass, which will be explained by factors like latitude, predation, and other features involved in natural selection. To be sure, some variability may remain unexplained, but this will not bother these scientists, from the outset convinced that sexual selection does not influence body size. Similarly, as CSR and other scientific approaches to religion discount *a priori* the role of revelation or religious experience in the formation of religious beliefs, it is unsurprising that they only uncover natural causes. CSR has enhanced our understanding of the cognitive basis of religious beliefs, but given its self-imposed methodological limitations, it is precluded from saying anything meaningful about the epistemic standing of religious beliefs.

Interestingly, phenomenological naturalness has been traditionally invoked as evidence for the truth of religious beliefs. Cicero, for example, argued that the existence of the gods is an idea natural to everyone, and engraven upon the mind (Cicero 45 BCE/1967, bk. 2). The theologian and reformer John Calvin wrote that humans have an innate sense of the divine that makes religious belief a deep-seated conviction in them; it can be perverted by sin and unwillingness, but it is nevertheless universally present:

> There is within the human mind, and indeed by natural instinct, an awareness of divinity… God himself has implanted in all men a certain understanding of his divine majesty…Yet there is…no nation so barbarous, no people so savage, that they have not a deep-seated conviction that there is a God…From this we conclude that it is not a doctrine that must be first learned in school, but one of which each of us is master from his mother's womb and which nature itself permits no one to forget. (Calvin 1559/1960, 43–46)

Following Calvin, Reformed epistemologists like Alvin Plantinga (2000) have argued that belief in God can be properly basic. A properly basic belief is a belief that is warranted even without relying on propositional arguments. For example, we believe in a basic way that the world is more than just a few minutes old, that other people have minds, and that our memories are generally reliable, even though there are no good arguments in support of these claims. Similarly, so it is argued, people can hold a spontaneous belief in God that is not based on arguments. Kelly James Clark and Justin Barrett (2010; 2011) have argued that

CSR provides empirical evidence for the proper basicality of religious beliefs, as it indicates that they are phenomenologically natural: they are held in the absence of arguments, and they emerge spontaneously as a result of normal cognitive processes. If religious beliefs are properly basic, we are justified in holding them without argument, just as we are justified in believing in the presence of other minds and the mind-independent external world. The Reformed case based on CSR indicates that phenomenological naturalness is not a promising starting point from which to debunk religious beliefs.

Turning to developmental naturalness, we have seen that some religious beliefs are the result of practiced natural skills. Some skilled practices are plausibly truth-tracking, whereas others are not. For example, a birder can accurately distinguish a buzzard from a sparrowhawk using visual clues such as differences between their silhouettes (e.g., tail shape) and their behavior (e.g., hovering). An astrologer who reads someone's horoscope relies on expertise about different signs and houses, and uses material tools like natal charts. The first practice is plausibly truth-tracking, but the second is not. The mere fact of being practiced natural says little about whether a particular skill provides one with justified beliefs. Without an in-depth examination of a practice, one cannot know whether it is truth-tracking. The question of whether religious practiced natural skills track truth cannot be answered as long as CSR attempts to reduce religious beliefs and practices to the nonreligious (e.g., byproducts of cognitive adaptations), without also acknowledging they are involved with a subject matter that is religious.

Conclusion

I have argued that NRT does not challenge the epistemic standing of religious beliefs. One reason why it does not is that CSR, like other scientific disciplines that study religion, has systematically attempted to reduce religion to other, nonreligious domains. For example, belief in an afterlife or reincarnation is regarded as a natural byproduct of an intuitive mind–body dualism. By not treating belief in God as belief in God, but as the output of a HADD or anthropomorphism, CSR leaves the epistemic standing of such beliefs unaddressed. Unlike natural histories of religion, CSR works descriptively, rather than normatively. However, if, as I have argued, religious practices and beliefs are not merely maturationally natural, but also practiced natural, we need to take them seriously as domains of skill and knowledge in their own right, and not just as phenomena that can be reduced to nonreligious domains.

Acknowledgments

Many thanks to Johan De Smedt and to the graduate students who participated in my cognitive science of religion seminar series at Oxford University for comments and discussion. This research is supported by a postdoctoral research fellowship (pf130006) from the British Academy.

References

Barrett, J.L. (2004). *Why Would Anyone Believe in God?* Lanham: AltaMira.

Barrett, J.L. (2010). The Relative Unnaturalness of Atheism: On Why Geertz and Markússon are Both Right and Wrong. *Religion* 40: 169–172.

Bering, J.M. and Bjorklund, D.F. (2004). The Natural Emergence of Reasoning about the Afterlife as a Developmental Regularity. *Developmental Psychology* 40: 217–233.

Bloom, P. (2007). Religion is Natural. *Developmental Science* 10: 147–151.

Boyer, P. (2002). *Religion Explained: The Evolutionary Origins of Religious Thought*. London: Vintage.

Buffon, comte de, G.-L.L. (1766). *Histoire naturelle générale et particulière. Avec la description du cabinet du roi*, vol. 14. Paris: Imprimerie royale.

Calvin, J. (1559/1960). *Institutes of the Christian Religion*, translated by F.L. Battles. Philadelphia: Westminster Press.

Carreiras, M., Seghier, M.L., Baquero, S., Estévez, A., Lozano, A., Devlin, J.T., and Price, C.J. (2009). An Anatomical Signature for Literacy. *Nature* 461: 983–986.

Cicero, M.T. (45 BCE/1967). The Nature of the Gods. In *Cicero in Twenty-Eight Volumes*, vol. 19, translated by H. Rackham, pp. 2–383. London: William Heineman.

Clark, K.J. and Barrett, J.L. (2010). Reformed Epistemology and the Cognitive Science of Religion. *Faith and Philosophy* 27: 174–189.

Clark, K.J. and Barrett, J.L. (2011). Reidian Religious Epistemology and the Cognitive Science of Religion. *Journal of the American Academy of Religion* 79: 639–675.

Comte, A. (1841). *Cours de philosophie positive: La partie historique de la philosophie sociale en tout ce qui concerne l'état théologique et l'état métaphysique*, vol. 5. Paris: Bachelier.

Cordry, B.S. (2011). A More Dangerous Enemy? Philo's "Confession" and Hume's Soft Atheism. *International Journal for Philosophy of Religion* 70: 61–83.

De Cruz, H. (In Press). The Relevance of Hume's Natural History of Religion for Cognitive Science of Religion. *Res Philosophica*.

De Cruz, H. and De Smedt, J. (2015). *A Natural History of Natural Theology: The Cognitive Science of Theology and Philosophy of Religion*. Cambridge, MA: MIT Press.

de Fontenelle, B.L.B. (1724/1824). De l'origine des fables. In *Oeuvres de Fontenelle, précédées d'une notice sur sa vie et ses ouvrages*, pp. 294–310. Paris: J. Pinard.

de Fontenelle, B.L.B. (1728). *Histoire des oracles*, revised edn. La Haye: Gosse & Neaulme.

Dennett, D.C. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. Oxford: Allen Lane.

Durkheim, E. (1915). *The Elementary Forms of the Religious Life: A Study in Religious Sociology*, translated by J.W. Swain. London: Allen & Unwin.

Evans-Pritchard, E.E. (1937/1965). *Witchcraft, Oracles and Magic Among the Azande*. Oxford: Clarendon Press.

Freud, S. (1927). *Die Zukunft einer Illusion*. Leipzig, Wien & Zürich: Internationaler Psychoanalytischer Verlag.

Gill, K.Z. and Purves, D. (2009). A Biological Rationale for Musical Scales. *PloS ONE* 4: e8144.

Goldman, A.I. (1999). A Priori Warrant and Naturalistic Epistemology. In *Philosophical Perspectives,* vol. 13: *Epistemology*, edited by J. Tomberlin, pp. 1–28. Oxford: Blackwell.

Gregory, B.S. (2006). The Other Confessional History: On Secular Bias in the Study of Religion. *History and Theory* 45: 132–149.

Guthrie, S. (1980). A Cognitive Theory of Religion. *Current Anthropology* 21: 181–203.

Guthrie, S. (1993). *Faces in the Clouds: A New Theory of Religion*. New York and Oxford: Oxford University Press.

Hardy, L. (2012). Hume's Defense of True Religion. In *The Persistence of the Sacred in Modern Thought*, edited by C.L. Firestone and N.A. Jacobs, pp. 251–272. Notre Dame: University of Notre Dame Press.

Hume, D. (1757). *The Natural History of Religion*. In *Four Dissertations*, pp. 1–117. London: A. Millar.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Kail, P.J.E. (2007). Understanding Hume's Natural History of Religion. *Philosophical Quarterly* 57: 190–211.

Kelemen, D. (2004). Are Children "Intuitive Theists"? Reasoning about Purpose and Design in Nature. *Psychological Science* 15: 295–301.

Luhrmann, T.M. (2012). *When God Talks Back: Understanding the American Evangelical Relationship with God*. New York: Vintage.

Malinowski, B. (1925/1992). *Magic, Science, and Religion and Other Essays*. Prospect Heights: Waveland Press.

McCauley, R.N. (2011). *Why Religion is Natural and Science is Not*. Oxford: Oxford University Press.

Plantinga, A. (2000). *Warranted Christian Belief*. New York: Oxford University Press.

Stark, R. (1999). Atheism, Faith, and the Social Scientific Study of Religion. *Journal of Contemporary Religion* 14: 41–62.

Stark, R. (2000). Religious Effects: In Praise of "Idealistic Humbug." *Review of Religious Research* 41: 289–310.

Winner, L. (2006). *Girl Meets God: On the Path to a Spiritual Life*. Milton Keynes: Authentic Media.

Wolfe, C.T. (2009). "Cabinet d'Histoire Naturelle," or: The Interplay of Nature and Artifice in Diderot's Naturalism. *Perspectives on Science* 17: 58–77.

Xenophanes. (6th century BCE/1903). Fragmente. In *Fragmente der Vorsokratiker*, edited by H. Diels, pp. 49–58. Berlin: Weidmannsche Buchhandlung.

Zuckerman, P. (2007). Atheism: Contemporary Numbers and Patterns. In *The Cambridge Companion to Atheism*, edited by M. Martin, pp. 47–65. Cambridge: Cambridge University Press.

34

Naturalism in Indian Philosophy

AMITA CHATTERJEE

Both "naturalism" and "Indian philosophy" have varied connotations and hence are in need of some tentative characterization. To begin with, let us take "an eschewal of the supernatural and a conviction that the methods of philosophy are continuous with those of the empirical sciences" (Ganeri 2012, 19) as the hallmarks of naturalism; while by "Indian philosophy" we shall refer to those systems of thought which were originated in India within the 1500 years of CE and their extensions.

Indian philosophy is a myth created by Indologists and orientalists. Just as there was no Indian nation before 1947, similarly there was no monolithic body of thought called "Indian philosophy" before the advent of orientalists. The geographical landmass known as Bhāratavarṣa has always been famous for its pluralism. It contained different systems of philosophy: some owing allegiance to the Vedic tenets (*āstika*-s), some going against them (*nāstika*-s). There were the Naiyāyika-s, the Vedāntins, the Cārvāka-s, and the Bauddhas, but one can find no trace of "Indian philosophy." During the colonial period, the orientalist myth was exploited by the nationalist leaders to create an Indian identity in order to boost the morale of the colonized people. "Indian philosophy" has thus become a way of clubbing together the classical systems of philosophy of the Bhāratavarṣa, mostly those which have been included in the famous compendium of Sāyana-Mādhava (1882), viz., *Sarva-darśana-saṃgraha.* We shall, therefore, confine our discussion to naturalist traits in those classical philosophical systems.

Since the study of Indian philosophical systems has been revived in Indian universities, academic philosophers have divided into three different groups. One group is engaged in the study of the classical Sanskrit philosophical texts, writing commentaries on them in Sanskrit, in some other Indian languages, and sometimes in English. Another group pursues Anglo-American or continental philosophies. The third group is involved in comparative philosophy, exploring and highlighting similarities and differences among different systems of thought with a view to initiating East–West dialogues. There are naturalists and non-naturalists in each group. Because of the

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

scientific temper of the day, contemporary philosophers seldom resort to anything supernatural to explain natural phenomena. The naturalist attitude is strong on the contemporary Indian philosophical scene, though there is much resistance to the project of naturalizing consciousness.

It is customary to distinguish Indian philosophy from the philosophy of the rest of the world by its spirituality, which enjoins liberation (*mokṣa*) as the *summum bonum* of human life. *Mokṣa* means going beyond the cycle of life and death, this world of experience, to attain a state which is free from all sufferings. If the whole of Indian philosophy is oriented toward attainment of liberation thus understood, then wouldn't a search for its naturalistic traits be an exercise in futility? Moreover, the Upaniṣadic literature, which is the storehouse of the seeds of all later philosophical ideas, claims two kinds of knowledge (*vidyā*): the higher (*parā*) and the lower (*aparā*). These two types of knowledge differ from each other in their objects, consequences, and methods of acquisition. *Parā-vidyā* is knowledge of the Brahman or the non-dual self, which leads to liberation (*niḥśreyasa mokṣa*), while *aparā-vidyā* is knowledge of empirical objects, means and ends, virtues and vices, which can at best lead to prosperity (*abhyudaya*) and heaven (*svarga*). With respect to methods of acquisition, there is opposition between two types of knowledge. *Parā* is said to be direct and intuitive (i.e., characterized by immediacy), while *aparā* has different means, including perception, inference, and so on. Hence, it is not possible to attain the higher knowledge by progressively mastering knowledge of empirical objects through logicoempirical methods. If there is an unbridgeable chasm between these two types of knowledge, how can one think of adopting a uniform method for the purpose of acquisition and explanation of their objects? Finding naturalism in Indian philosophy, therefore, seems highly improbable.

Even if these objections are legitimate, Indian philosophy, again, does not have a monolithic structure. Alongside spiritual systems (*adhyātma-vidyā*) like the Vedānta and the Sāṃkhya-Yoga, there developed an analytical (*ānvīkṣikī*) tradition of enquiry within the orthodox fold, mainly represented by the Nyāya-Vaiśeṣika system, which employed a common method of enquiry for both science and philosophy, prescribing a way to liberation by removal of ignorance of the real nature of things contained in our world of experience; Nyāya-Vaiśeṣika can take naturalism in its stride. In addition, even within the spiritual (*ādhyātmika*) tradition, there are philosophers who have shown the compatibility between the higher and the lower knowledge from a top-down perspective. For them, accommodating naturalism is not an impossible task.

Naturalism and Nature

"Naturalism" shares the ambiguity of its root "nature." It has often been pointed out that, in a particular philosophical context, the meaning of "nature" can be best determined by raising the question, "what is nature (or the natural) being contrasted with in this context?" – artificial, conventional, miraculous, transcendental, spiritual, ritualistic, or what? (Hepburn 1967) We shall follow this pragmatic advice in our understanding of "naturalism."

Naturalism as a brand of philosophy need not be a global theory. As with realism, naturalism admits of many different kinds: ontological, methodological, semantic, linguistic, moral, and aesthetic. Proponents of one may not subscribe to another. Ontological

naturalism prohibits the entry of the supernatural into the realm of the so-called natural.[1] Methodological naturalism opposes any form of a priorism or transcendentalism (in the Kantian sense) in theorizing. Semantic naturalism refuses to admit empirically unanalyzable theoretical entities. Linguistic naturalism rejects conventionalism, holding that names have a kind of natural attachment to their objects. Moral naturalism holds that good and bad, for example, can be explained in terms of natural properties such as human preferences. Finally, aesthetic naturalism concedes no non-natural criterion of aesthetic judgment and appreciation.

Is there anything in common among the different kinds of naturalism? Each seems to agree that its domain is amenable to scientific explanation through empirical investigation and causal enquiry. Naturalism, on this count, is more methodological than ontological (Danto 1967). Adopting methodological naturalism leaves one free to draw the line between natural and supernatural in accordance with one's discipline, fiat, and theoretic and pretheoretic commitments. For example, an ontological naturalist may seek an understanding of humans as part of nature and yet place freedom, for example, outside the realm of nature. As a methodological naturalist, one might be a reductionist or eliminativist about distinctive human traits and abilities, all the while embracing some kind of nonreductive dualism. P.F. Strawson (1985) suggests a more systematic classification of ontological naturalism into hard/reductive and soft/liberal. Hard naturalists view human endowments as objective parts of nature, which are therefore characterized by unity and ahistoricity. Soft naturalists, on the other hand, favor the appropriate relativization and history-specific assessments of idiosyncratic human nature within a stable, overarching framework.

How is naturalism understood in various forms of so-called Indian philosophy? The Sanskrit word for "nature" (*prakṛti*) has at least three different senses relevant to the present context. First, *prakṛti* can mean the collection of all of the objects in the world; second, it can be synonymous with the essential nature (*svabhāva*) of anything, including its dispositional qualities; third, it can mean the original source of the material world, the locus and object of our experience. Which type of ontological naturalism a philosophical system upholds – hard or soft – depends on the sense in which it takes *prakṛti* (Chatterjee 2012).

Let us begin with *prakṛti* as the collection of all of the objects in the world. Indian philosophical systems, like other philosophical systems, are rarely in agreement on what there is in the world. Take, for instance, consciousness. In the Advaita Vedānta system (Dharmaraja 1970), Self or Brahman, which is self-luminous consciousness, stands apart from the empirical world, so an Advaita Vedāntin is a non-naturalist about consciousness. According to the materialist Cārvāka philosophers, consciousness is an emergent property of material elements and there is no other realm of conscious objects than this material world. A Cārvāka is, therefore, a hard or reductive naturalist about consciousness. Nyāya-Vaiśeṣika philosophers, for a third view, hold that the self is an immaterial substance, and that consciousness is an accidental property of an embodied self; in its disembodied/liberated state, an individual self is unconscious. Consciousness, on this view, depends on embodiment.

1 The Cārvāka-s, the hardcore naturalists, for example, admit consciousness in their metaphysics, but not soul as the disembodied substratum of consciousness; they admit four basic material elements as the cause of the world, but do not feel the necessity of accommodating God in that role. The mitigated naturalists, Naiyāyika-s, on the other hand, admit consciousness as well as embodied souls as the substratum of consciousness and accommodate the creator God as the efficient cause of the world over and above its material causes, by propounding a different scheme of causation.

Since philosophers belonging to the Nyāya-Vaiśeṣika school thus accommodate consciousness in the realm of experience, they can be considered soft or liberal naturalists.

Prakṛti as the essential nature (*svabhāva*) of a thing leads to a view known as *svabhāvavāda*, which literally means "the view of individual nature/essence." *Svabhāvavāda* reminds us of Aristotle's statement, "Of things that exist, some exist by nature, some from other causes" (*Physics* 192b9–192b11). An oft-quoted verse depicting this view says, "the fire is hot, the water cold, refreshing and cool is the morning breeze, by what came all this variety? From individual nature was it born" (Bhattacharya 2002, 78). *Svabhāvavāda* does not allow one to raise questions, such as "Why is a thorn sharp?" or "Why is a rose fragrant?" No further componential or causal analysis is possible. The buck stops at "individual nature." Reductionist Cārvāka-s and nihilist Ājīvika-s, for example, do not believe in any causal story of the origin of this world. This is probably the strongest form of ontological naturalism. We shall come back to this view later.

Finally, consider the third sense of *prakṛti* as the source of the material world. Like some early Greek philosophers, some ancient Indian philosophers were also natural scientists, and attempted to trace the origin of the world to some natural stuff. The Cārvāka-s held that four basic physical elements (*bhūta*-s), viz., earth (*kṣiti*), water (*ap*), fire (*tejas*), and air (*marut*), combined in different proportions to give rise to the living and the nonliving world. Nyāya-Vaiśeṣika-s, Buddhists, and Jaina-s proposed different versions of atomic theory (*paramāṇuvāda*), while Sāṃkhya-Yoga philosophers maintained that the world is a transformation of the Ur-Nature (*mūla-prakṛti*), and the monist Vedāntins traced the origin of the apparent world to ultimate reality or Brahman. Two parameters help us to determine which of these systems are naturalist, to what degree they are naturalist, and what kind of naturalist they are: the nature of real entities admitted in a system and the nature of causal connections obtaining among these entities. Interestingly, if God is bound by the causal laws that govern the world, then God is immanent in nature, thus blurring the line between the natural and the supernatural. A discussion of some of these classical Indian philosophical theories will show how ontological naturalism varies from system to system.

Indian Concepts of Nature

Atomism

Nyāya-Vaiśeṣika philosophers (Praśastapāda 1930) were the most thorough of the Indian atomists. Buddhist and Jaina philosophers also added interesting dimensions to atomism. However, atomists across systems arrived at their conclusions on the basis of meticulous observation and logical speculation. In absence of experimental evidence and lacking an adequate theoretical framework, they could not anticipate the results of modern atomic physics or quantum mechanics. But their insights regarding the composition of atoms led to the remarkable development of applied chemistry and applied medicines in ancient India, and the associated debates about the nature of causation provided us with the philosophical foundations of scientific enquiry.

Nyāya-Vaiśeṣika philosophers were realists and pluralists. They admitted in their ontology nine kinds of substance: earth (*kṣiti*), water (*ap*), fire (*tejas*), air (*marut*), *ākāśa* (the substratum of sound), space (*dik*), time (*kāla*), soul (*ātman*), and an inner sense organ called mind (*manas*). Of these, the first four and the last one are of the smallest magnitude

(*aṇu-parimāṇa*) and the other four are all-pervasive (*parama-mahat-parimāṇa*). All atoms, the indivisible simples of the world of experience, are eternal (*nitya*), as are *ākāśa*, space, time, and soul. But the composite wholes arising out of the combination of atoms are non-eternal (*anitya*), medium-sized (*madhyama-parimāṇa*), and eventually destroyed. Atoms of all four elements are spherical and quantitatively identical, though atoms of each type possess a specific attribute, which accounts for their qualitative differences.

Nyāya-Vaiśeṣika-s based their arguments for atoms on two presuppositions: (1) a part of a whole is always smaller in size than a whole and (2) the part of the smallest visible composite thing is imperceptible. According to them, the smallest motes swimming in a sunbeam are the smallest visible things, and hence these must also possess parts. But the process of dividing composite things into their parts cannot go on indefinitely. So there must be partless, indivisible, imperceptible things. If an infinite division of composite things were possible, a mustard seed and a mountain would be equal in size (as both of them would consist of an infinite number of parts, and infinities are all equal).[2] Division must stop at atoms, because a process ending in nothing is as absurd as digging a hole in empty space; additionally, this would imply that the world can be created out of nothing, and Indian philosophers generally opposed the idea of creation *ex nihilo*.

Atoms of four different kinds possessing four different specific qualities (e.g., odor of earth atoms, taste of water atoms, color of fire atoms, and touch of air atoms) combine in different proportions to give rise to this world of multifarious objects. The process of combination needs to be described in detail, not only because there are internal debates on this point, but also because it will reveal some of the peculiar features of Indian naturalism.

Two atoms of the same type combine to form a dyad (*dvaṇuka*), and three dyads of the same type form a triad (*trasareṇu*), the smallest visible object. From triads onward, the possibility of heterogeneous combination of different elements giving rise to different structural arrangements is admitted. Dyads are called *ārambhaka*-s, the primary building blocks in the system, because the process of creation starts from this stage, single atoms being eternal. Once contact between dyads is destroyed, the world comes to an end (*pralaya*) and, later, a new cycle of creation begins, with the formation of new dyads. If two atoms of different types could combine – say, for example, an earth atom with a fire atom – the resultant could not form a natural kind. Nyāya- Vaiśeṣika-s maintain that both atoms and dyads are imperceptible; visibility arises only at the level of triads. Since single atoms are of imperceptible magnitudes, so are dyads. Though dyads form a new whole and are not just a combination of parts, these wholes lack a perceptible dimension, because any quality of a part, including its magnitude, produces the same quality in the whole, to a greater degree. Thus, dyads have an even more imperceptible magnitude than atoms. How can one beat this logic at the level of the triad? How is a whole of perceptible magnitude possible? The answer is that the magnitude of a compound may be due to either the grossness of its components, the looseness of the conjunction of parts, or the plurality of their numbers. Atomists deny the possibility of the first alternative. The second is also ruled out, because they do not admit any interstice between two atoms. So, the perceptible magnitude of a triad must be due to the number of its components: the plurality of the number of components is the cause of the grossness of a compound. According to Sanskrit grammar, the first plural number is three. So, visibility is generated with a combination of three dyads.

2 This view was developed long before Cantor proved his theory of transfinite cardinals.

However, to rule out the generation of visibility by the combination of three atoms, Nyāya-Vaiśeṣika-s come up with another principle, which states that a product of a perceptible magnitude can be generated only by a product of smaller magnitude. Though atoms are the smallest units, they are not products. The smallest product comes into existence with a dyad, and, hence, only three dyads, not three atoms, can give rise to the perceptible dimension. These philosophers are liberal naturalists who accommodate numbers in their capacious worldview not merely as nomological danglers, but as having causal effects on the physical world.

The atomic theory of any system is often related to its theory of causation. Nyāya-Vaiśeṣika philosophers were *asatkāryavādin*-s, meaning they did not admit the existence of an effect in its cause prior to its production. Rather, they held that every effect is a new beginning. There is a prior absence of an effect in its cause before the effect is produced. Therefore, a cause is an invariable, unconditional, immediate antecedent of an effect, and an effect is the countercorrelative of a prior absence. Causes are classified in this system first into two types: common (*sādhāraṇa*) and uncommon (*asādhāraṇa*). A common cause is necessary for effectuation as such, while an uncommon cause produces a particular effect, being its immediate, invariable, and unconditional antecedent. Common causes include space, time, the accumulated merits and demerits of individual agents (*adṛṣṭa*: literally, "unseen"), God, knowledge, desire and will (*prayatna*) of God, and prior absence (*prāgabhāva*). Uncommon causes are divided into three classes: (1) inherent (*samavāyi*), (2) noninherent (*asamavāyi*), and (3) efficient (*nimitta*). For example, atoms are the world's inherent causes, conjunctions of atoms are its noninherent causes, and God and *adṛṣṭa* are its efficient causes. Does the admission of unseen forces and divine intervention to impart the motion in naturally inert atoms compromise this brand of naturalism? We leave this as an open question at this juncture, and shall come back to it in our discussion of moral naturalism.

Another interesting feature of Nyāya-Vaiśeṣika atomism is that *manas* (the inner sense organ) is also said to be atomic in magnitude and corporeal in nature, though it lacks a special quality. Because of its nature, *manas* can easily build a bridge between knowable objects in the physical world and the knower self without encountering the problems of Cartesian dualism.

Realist schools of Buddhism (viz., the Vaibhāṣika and the Sautrāntika) and the Jaina-s also propounded atomism. Unlike Nyāya-Vaiśeṣika-s, these heterodox schools offered accounts of combinations of atoms in terms of natural forces. They did not fall back on God's will and unseen forces to explain the process of combination. Jaina-s (Umāsvāti 2007) also speculated on the nature of atomic motion and formulated rules of atomic combination. Śubhagupta, a later Vaibhāṣika, offered an alternative account of the combination of atoms, in which two atoms come close to each other because of their inherent potency (*dravyaśakti*), though they are not actually conjoined (Matilal 1986). The aggregate thus formed combines again to give rise to varied composite objects of the world. However, not all atoms are equally potent, and some never become a part of an aggregate, because of their insufficient bonding power. Atoms, when bonded together, undergo a transformation due to mutual influence, and novel properties emerge in the aggregate that were not present in the single atoms. For example, carbon compounds, when transformed into diamonds, become too hard to be disintegrated (Gangopadhyay 2008).

Some Buddhists, we should note, are also ontological naturalists in a different sense. The Buddha did not admit God as the creator of the world, nor was he a message-bearer of an

omniscient God, nor a representative of any revealed scripture. He repeatedly told his disciples that he had attained enlightenment through his own effort and that it is possible for every human being to attain it irrespective of caste, creed, or gender. He did not want his disciples to follow his teachings blindly, but urged them to examine his sayings in the light of reason. The Buddha did not believe in performing miracles and despised those who deceived people and claimed that they possessed supernormal powers. Yet Buddhist philosophy has been criticized as non-naturalist on three grounds. First, the admission of post-mortem existence seems contra naturalism; second, the Buddha often spoke of gods and superhuman beings (mentioned in Nikāya texts and *Jātaka* stories); and third, he preached the virtue of meditation and the possibility of developing supernormal powers through meditation. If the Buddha was a naturalist, he wasn't the typical naturalist.

The Buddha believed in the ubiquitous reign of causation, which, in the moral realm, takes the form of the law of *karma*; the thesis of rebirth follows logically from this law. As one reaps the consequences of one's own actions, and as the results of one's actions seldom bear fruit within a single lifetime, Buddhists affirm rebirth.[3] If the phenomenon of rebirth is accepted, then Buddhists must admit some extraterrestrial realms where beings can be reborn in accordance with their past actions (*karma*-s). Buddhists admit 31 worlds arranged in the order of available comfort in which one can be reborn. The gods (*deva*-s) that the Buddha talked about are not supernatural beings. They are inhabitants of other worlds and attain their status according to their previous actions. Finally, the Buddha did prescribe two kinds of meditative reflection as useful means of training minds: mindfulness and concentration. However, these are neither mystical processes nor forms of silent prayer. These are the means that transform an ordinary person into a super person and endow one with perfection. Buddhism thus retains ontological and moral naturalism in a qualified sense.

Hard Naturalism

According to Buddhists, Jaina-s, and Nyāya-Vaiśeṣika-s, the natural world is governed by causal laws. And yet they accepted the phenomenon of mental causation and also causation across different worlds. The reductive naturalist Cārvāka-s, on the other hand, believed in the causal closure of the physical universe arising out of the combination of the four basic material elements. They rejected imperceptible things and therefore excluded atoms, *ākāśa* (the fifth material element admitted in the orthodox systems), God, the soul, and all other non-natural forces from their ontology. Since they viewed consciousness as an emergent quality of matter, the conscious living body (the reference of "I") can easily enter into causal relationships with other natural objects. Another group of extreme naturalists (*svabhāva-vādin*-s) refused to admit any causal foundation of the natural world. They held that objects possess some qualities naturally. The nature of composite objects does not causally depend on the nature of their components. Consequently, they believed in the fortuitous generation of events (*ākasmikatāvāda*). Causal relations, they held, are supposed to involve necessity; but necessity is imperceptible, and whatever is not perceptible cannot be inferred or established by any other means. Udayana (1996), a 10th-century Naiyāyika, exposed the hollowness of this view and the meaninglessness of the thesis as a

3 Rebirth should not be confused with reincarnation. Reincarnation assumes a transmigrating soul that passes from one life to another. But Buddhists do not admit a transmigrating permanent soul.

whole. The crux of Udayana's arguments is as follows. Every event must have a cause, because every event without exception has "conditional" (*sāpekṣa*) existence, this in turn because it has "occasional" (*kādācitka*) existence; that is, it occurs at a certain time. An eternal entity is always existent and a fictitious entity does not exist at any time (as these are not characterized by occasional existence, they are not caused). But each event in the natural world happens at a particular moment of time, because it is due to a specific cause. The extreme naturalists affirm, however, that an event need not originate from a cause; it may come into being fortuitously. Even the occasional origination of an event is due to the nature of the event and has nothing to do with its cause. Udayana next proceeded to demolish the thesis on the basis of an etymological analysis of the word *akasmāt* (without cause). He listed five possible meanings of this word, viz.:

(1) An effect does not originate from a cause.
(2) An effect does not arise at all.
(3) An effect is self-caused; it is not determined by any external condition.
(4) An effect is generated by an unreal cause.
(5) The occurrence of an effect is not determined by its cause but by its own nature (*svabhāva*).

Udayana (1996) objected to each of these formulations. If an effect was not dependent on its cause for its existence, then it could have occurred at any time, in fact at all times, and thus would lose its occasional nature. Every effect has a temporal limit fixed by its cause, prior to which it cannot exist. (2) runs contrary to our perception of the occurrence of an event at a particular spatiotemporal location. (3) is unacceptable because the same thing cannot be both a cause and an effect at the same time in respect to the same set of conditions, and because it is not possible for anything to exist before its origination. (4) can be rejected because no unreal thing can ever enter into a causal process. And (5) leaves us mystified, because the proponents of the fortuitous-generation thesis have not specified what this nature is by virtue of which an effect can occur without its cause. So we wonder, is this nature different from or the same as the effect? On the first alternative, the principle of causality is reestablished, while the second alternative is unintelligible. For, if this nature were the same as the effect, the effect would be continuously generated. The doctrine of the fortuitous generation of events is therefore absurd and stands in the way of progress of scientific knowledge.

Protonaturalism

The Sāmkhya view of nature (Iswarakṛṣṇa 1972) is, at first sight, protonaturalist (primitive but not full-fledged). The Sāṃkhya and its associated system, Yoga, admit 25 principles into their ontology. Of these, *Puruṣa* or self is an autonomous witness of the natural world but does not play any part in it. All other principles out of which the natural world evolves are evolutes of the all-pervasive, ever-dynamic Ur-Nature (*Mūla-Prakṛti*, or *Prakṛti* for short). This Ur-Nature comprises three constituents, called the *guṇa*-s, viz., *sattva*, *rajas*, and *tamas*. *Sattva* is light and illuminating (i.e., it naturally reveals objects), rajas is restless and active, and *tamas* is heavy and restraining, and covers the nature of objects. The three *guṇa*-s are sometimes interpreted as form, energy,

and mass (Seal 1958), respectively, and sometimes again as pleasure, pain, and indifference (Bhattacharya 1956; Mohanty 1992). They are present in all objects of the world in different proportions and are responsible for our varied experience. Ur-Nature is not perceptible. One can infer its existence from its perceptible evolutes. As all objects of the world affect us either as pleasurable or as painful or as indifferent, depending on the dominant presence of one of the *guṇa*-s, we can infer that Ur-Nature is also characterized by these qualities. But, to establish this, Sāṃkhya philosophers require a commensurate theory of causation to show that cause and effect are identical. Accordingly, they offer their distinctive theory of causation, *satkāryavāda*, which states that an effect exists in its cause prior to its production in a latent form.

Sāṃkhya philosophers advance five main arguments in support of their theory of causation (Chatterjee 2006). (1) What is nonexistent can never be produced. No agency can turn the nonexistent into the existent. So, if the effect were nonexistent in the material cause before the causal operation, then it would never be produced. (2) A cause produces an effect only by being related to it. But for a relation to obtain requires at least two relata. Hence, an effect must preexist in its cause. (3) If the effect could arise without being related to its cause, then any cause could give rise to any effect. If there were no definite relation between threads and cloth, why couldn't a pot arise from threads? (4) For a cause to produce an effect, it must have the adequate potency; otherwise, the desired effect cannot be produced. Since oil seeds possess adequate potency, oil can be produced out of these seeds (but not out of sand). Sāṃkhya philosophers concede this point and maintain that causal operations enable a potent cause to manifest its latent effect, and that this causal efficiency is nothing other than the existence of the effect in the material cause in a latent form. (5) The effect exists in its material cause because cause and effect are essentially the same, but different in form. Since the cause is existent, the effect must also exist. These arguments imply the existence of Ur-Nature and also show against the Vedāntins that the natural world is a real transformation of Ur-Nature and not mere illusory appearance.

Let us now look at the Sāṃkhya account of the evolution of the world from Ur-Nature. The first evolute of Ur-Nature is the *mahat-tattva* (the Great Principle, the Cosmic Intelligence, or *Buddhi*). From this emerges I-consciousness (*ahaṃkāra*). From the *sattva* aspect of I-consciousness evolve five organs of knowledge (eye, ear, nose, tongue, and skin), five motor organs (speech, hands, feet, and reproductive and excretory organs), and *manas* (sometimes translated "mind"); from the *tamas* aspect of I-consciousness emerge five subtle elements (*pañca-tanmātra*: sound, touch, color, taste, and smell). The five subtle elements give rise to five gross elements (*ākāśa*, air, fire, water, and earth). Our world of experience emerges from these gross elements combined in different proportions. The process of evolution and dissolution goes on cyclically.

Ur-Nature, according to the system, is always in transformation. Before the beginning of the creation or the empirical manifestation of Ur-Nature, there is a homogeneous transformation (*sadṛśa-pariṇāma*) of the constituent principles; that is, *sattva* transforms into *sattva*, *rajas* into *rajas*, and *tamas* into *tamas*. So, Ur-Nature remains in a stable state. At the time of the creation of the world, the stability of Ur-Nature is disturbed due to its close proximity with the self (*Puruṣa*), an independent co-eternal reality (as a piece of iron in the proximity of a magnet), and the process of heterogeneous transformation begins. The active principle, *rajas*, becomes dominant and activates the other two principles. Then these three principles combine with one another in different proportions and the manifold world comes into existence.

The process of the evolution of Ur-Nature takes place, says the tradition, for the enjoyment and liberation of the self. Though the self stands outside the process of evolution, he conflates his identity with the first evolute or *buddhi*, the latter being transparent due to the preponderance of the *sattva* element in it. As a result, the self seems able to experience enjoyment and suffering. When the self attains discriminatory knowledge, he overcomes this confusion, realizes his own nature, and is liberated. The Sāṃkhya theory of evolution is therefore described as teleological. A careful reading of this analysis immediately reveals that there are many gaps in this evolutionary narrative. It is not clear why Ur-Nature should try to entrap and bind the self, nor why confusion in the self begins at all. Hence, Sāṃkhya philosophers maintain that the enjoyment and the liberation of the self are also appearances, because the self is eternally free. On this reading, however, teleology loses its force. Mohanty (1992) thinks that is perhaps inevitable, because the Sāṃkhya teleology had always been protonaturalistic. Mohanty offers two examples from the literature to prove his point. Just as nonsentient cow milk flows naturally for the nourishment of the young calf and nonsentient rain clouds naturally yield rain for the sustenance of life on earth, so Ur-Nature binds the self for the latter's enjoyment and liberation. The directedness of Ur-Nature and its evolutes toward the satisfaction of another's need is thus entirely natural and unconscious. Dasgupta (1987) conjectures that the Sāmkhya system upholds dualism to explain the normative and spiritual aspect of a conscious self (which has always been found to be a stumbling block for naturalism as such). By admitting dualism, it is possible to show that an embodied individual is a part of the natural process of evolution. But the introduction of conscious self and personality marks a new kind of progress, which is based on but is not exhausted by the natural process of evolution. *Puruṣa*, therefore, does not participate in the process of evolution. However, that does not make empirical consciousness naturalistically inexplicable. In the world process, *buddhi* plays a conscious role, reflecting the pure consciousness of self, just as the moon illumines the world by borrowing the reflected light of the sun.

Methodological Naturalism

Methodological naturalism (Kornblith 1985; Papineau 2009) admits continuity between empirical enquiry in science and theorizing in philosophy with respect to the method(s) adopted or the results arrived at through scientific enquiry, or both. Methods continuity prescribes that theorists, while building their theories, must emulate a scientific way of understanding the world. Methods continuity may be strong or weak. The strong view holds that all theories should be developed following the methods of physics, the so-called "ideal science," while the weak view permits theories compatible with different methods adopted in special sciences. Result continuity, on the other hand, implies that philosophical theories be warranted or supported by results of the relevant sciences. Goldman, for example, takes into account the results of cognitive psychology regarding how people actually form beliefs to answer the question how people ought to form beliefs.

Most post-Kantian philosophers distinguished philosophical and scientific endeavors in terms of the methods used and contents discussed. The scientific approach is said to be descriptive and scientific methods empirical, while the philosophical approach is normative and philosophical method *a priori* and logical. The 20th-century linguistic philosophers

maintained that science is a collection of descriptive propositions about the world but philosophy is an act. Wittgenstein declared in the *Tractatus* (1961)[4] that the task of philosophy is to give meaning to scientific concepts and that philosophy is not on a par with sciences: it is to be placed either above or below sciences. Contra these views, methodological naturalists "see philosophy and science as engaged in essentially the same enterprise, pursuing similar ends and using similar methods" (Papineau 2009).

In the Indian theoretical tradition, putting science and philosophy in the same continuum never appeared a problem, because there were no strict disciplinary barriers. Progress in human knowledge thrived on cumulative efforts. So music, mathematics, and morals or physics, philosophy, physiology, and psychology were never placed in hermetically sealed compartments. Different philosophical systems combined the metaphysic of the transcendent with the logic of the mundane. That is why there were scientific elements in philosophy and elements of transempirical philosophy in the sciences. The systems which provided the ratiocinative methods of scientific enquiry also facilitated the process of self-realization, culminating in liberation through discourses on the nature of reality. B.N. Seal therefore traced the positive sciences of the ancient Hindus, comprising discussions on cosmology, astronomy, physics, chemistry, biology, psychology, and so on, within the philosophical discourse.

In this context, it will be relevant to point out that Naiyāyika-s in the beginning of the first millennium after Christ prescribed a general method of enquiry which was used by both scientists and philosophers alike. Neo-Nyāya philosophers (Gangeśācārya 1974) in the 13th century developed a quasi-formal language that was adopted by most participating in any theoretical discourse from the 15th century onward, be it grammar or mathematics, law or literary theory. The Nyāya method of enquiry has four main steps. The first is to provide an enumeration (*uddeśa*) of the topics under discussion. The second is to offer a (real, as opposed to nominal) definition (*lakṣaṇa*) of the subject under consideration, specifying its distinguishing mark as well as the scope of its application. The third step is an examination (*parīkṣā*) of the definition, and the fourth is verification (*nirṇaya*). Enumeration sometimes includes classification (*vibhāga*); however, in general, classification comes after definition. Any truth reached by this procedure is raised to the status of an established theory (*siddhānta*). Thus, in both method and content, philosophy and the theoretical sciences coincided to a large extent.

Methodological naturalism can be traced very clearly in epistemological discussions, especially in the Nyāya epistemology, where common sense, science, logic, and scriptures are all found to converge. We shall therefore take up next the discussion of the Nyāya epistemology, which will be interpreted as a kind of naturalist epistemology.

Naturalism in Nyāya Epistemology

Epistemology has always been considered a normative discipline par excellence and much of Western analytic epistemology developed around the concept of epistemic justification. The epistemologists' preoccupation with the formulation of principles of epistemic appraisal and grading of intellectual acts along the epistemic dimension were all geared to meet the

4 "4.111 Philosophy is not one of the natural sciences. (The word 'philosophy' must mean something whose place is above or below the natural sciences, not beside them.)" (Wittgenstein 1961).

skeptical challenges. As effective means of silencing the skeptics, they adopted three main strategies. (1) They granted autonomy to epistemology, since epistemology is supposed to provide the basis for all human scientific endeavor. (2) They tried to ground our knowledge of the external world on the subject's knowledge of inner experience, which led to the acceptance of an internalist theory of justification. (3) They declared all epistemic norms of justification *a priori* in nature. They further maintained that providing causal explanations is no part of epistemology (Chisholm 1992). Causal questions and justificatory questions are distinct. So, to judge whether a person's belief that *p* counts as knowledge that *p*, it is sufficient to find out if *p* is connected in the right way to other propositions that are believed, and the rightness of the connection is determined in terms of logic (Kitcher 1983). It is irrelevant to probe the origin of *p*. Though naturalized epistemologists formed a heterogeneous group, they contested the supposedly "Archimedean standpoint": the scheme of internalist justification and apsychologism of the traditional epistemology.

Naturalized epistemology was introduced to the contemporary philosophical scene by W.V.O. Quine in his famous essay, "Epistemology Naturalized" (1969). Even a cursory reading of the text reveals that Quine was exhorting epistemologists to give up the normative project of epistemology altogether. He was urging a replacement of traditional analytic epistemology by naturalized epistemology, a law-based predictive explanatory theory like any other theory within the empirical sciences. Thus, he robbed epistemology of its traditional role of acting as the arbiter and the arbitrator of foundational scientific disputes. Quine realized later that his earlier view was too radical and therefore hinted at a moderate naturalized epistemology with the following three features: (1) epistemology must be continuous with science; (2) causal questions must form part of epistemology (i.e., epistemologists should take stock of the psychological conditions of cognition); and (3) though traditional models of justification are unacceptable, epistemologists may resort to a scientific model of justification, combining features of both foundationalism and coherentism (Barrett and Gibson 1990). Quine thus reinstated justification, but of a different kind: some consider justification in terms of causally reliable processes of belief generation to be adequate, while others claim that justification supervenes on natural facts.

The most general arguments in favor of the claim that most Indian *pramāṇa theories* (epistemic theories) are naturalistic are as follows. Each develops its respective theory of veridical cognition and/or knowledge (*pramā*) in response to skeptical threats. Despite differing metaphysics, most attempt to explain cognition with reference to a psychocausal chain. As far as knowledge of the empirical world is concerned, all except the radical skeptics admit the primacy of perception and thus provide the systems of epistemology with a strong empirical foundation. Indian traditions, in general, sustain a methodological continuity between science and philosophy. Indian philosophers belonging to the *ānvīkṣikī* tradition did not feel any need for the *a priori*/*a posteriori* distinction, nor did their theory depend on necessity/possibility or analytic/synthetic distinctions. As a result, they could easily commute between the realm of the normative and that of the descriptive. All these are indicative of a moderate methodological naturalism; however, in the absence of any special scientific domain, they do not lend support to a radical replacement naturalism.[5]

5 Radical replacement naturalists do away with normative justification theories and try to replace them with empirical and descriptive scientific explanation. Quine at one stage upheld that epistemology should be replaced by psychology.

Of all Indian epistemic systems, moderate naturalist epistemology's features can be traced most systematically to Nyāya epistemology (*Pramāṇavāda*). In the early 2000s, many scholars attempted to show that the Naiyāyika-s had a concept of justification analogous to that of the classical epistemologists (especially the justified-true-belief theorists). To establish this position, however, they had to read between the lines of the Nyāya texts and reconstruct the Nyāya position in a rather tortuous manner. But a nontortuous reading of the Nyāya texts appears to support the thesis that the Naiyāyika-s were moderate naturalists.

The Naiyāyika-s could respond to skeptical challenges in two steps. First, they enumerated and explained a set of reliable processes that are more likely to generate true beliefs than false beliefs, and second they attempted to ratify those processes of belief acquisition. In Nyāya epistemology, four kinds of cognition, viz., perceptual (*pratyakṣa*), inferential (*anumiti*), comparison-based (upamiti), and verbal (*śābda*), are admitted, and correspondingly there are four accredited means of acquiring veridical cognition (*pramāṇa*), viz., perception (*pratyakṣa*), inference (*anumāna*), comparison (*upamāna*), and authority (*śabda*). The Naiyāyika-s decided the number of the accredited means by empirically observing the effectiveness and reliability of the respective means of generating true beliefs. According to them, these processes generate true beliefs only when accompanied by a genuine excellence or epistemic virtue (*guṇa*). Such epistemic virtues differ in producing different types of true belief. In the case of perception, the epistemic virtue is identified with the relation of the sense organ with the object characterized by the property which figures as the qualifier in the perceptual cognition. For example, when one perceives a terracotta jar as red, one's visual sense organ stands in the appropriate relation to the object of perception, a jar, and apprehends the property redness which characterizes the jar in question and thus gives rise to a veridical perception. If one perceives a red jar as blue, then it is an erroneous perception. In a veridical inferential cognition, the mark, which is invariably concomitant with the thing to be inferred, must be known to be present in the locus of the inference. For instance, when someone correctly infers fire from smoke on a distant hill, it is known to him that smoke, which is invariably concomitant with fire, is present on the hill. In cases of comparison-based cognition, knowledge of similarity is the excellence; for example, when someone rightly identifies an unknown animal as a bison on seeing its similarity to a cow, having came to know from an expert's statement that a bison is similar to a cow. Finally, in cases of verbal knowledge, the speaker's veridical cognition of the state of affairs described by the sentence uttered is the excellence; for example, when a doctor declares that a patient is suffering from malaria on the basis of his expertise and veridical cognition of the state of affairs. A false belief results from the presence of some defect (*doṣa*), and not merely from absence of the required virtue. One may perceive a white shell as yellow because one is suffering from jaundice or because of the yellow-tinted light in the room. Defects differ in each instance of false belief. Naiyāyika-s therefore maintain that a false belief is caused by a defect and a true belief is caused by a virtue. This principle holds in all cases and all types of belief – commonsensical, scientific, or philosophical. However, though a defective process usually generates a false belief and a meritorious process generates a true belief, some beliefs may be true by fluke, in spite of being produced by a defective process. Suppose someone wrongly perceives dust clouds as smoke and argues, "The hill is on fire, because it has smoke on it." If, unbeknownst to him, the hill actually is on fire, then this argument yields a veridical conclusion though its ground is defective. Thus, Naiyāyika-s uphold the principle that "If there is a false belief, then there must be a defect in the generating process," but not its

converse: "If there is a defect in the generating process, then it produces a false belief" (defects are necessary but not sufficient for error). Such considerations have led scholars like Sibajiban Bhattacharyya to declare that Naiyāyika-s did not mean that veridical cognition (*pramā*) entails a justified true belief. Others, including J.N. Mohanty (1992; 2001), contest this view. They reconstruct the notion of *pramā* as a justified true belief and include cases similar to the one just mentioned in the list of the Gettier-type counterexamples, thereby attempting to accommodate *pramāṇa* theories within the framework of traditional epistemology. Even then, one may hold that Nyāya epistemology remains within the naturalistic framework if, as modern interpreters of the Nyāya epistemology assert, there is no distinction between the causal and justification conditions of a piece of cognition.

The naturalist stance of Nyāya becomes clearer if we consider the Naiyāyika view of apprehending cognition. A veridical cognition, says Vācaspati Miśra, cannot reveal its own truth, nor can it be grasped in after-perception; it must be apprehended by a subsequent inference following from volition leading to successful activity. Consider a typical example. Suppose a thirsty traveler perceives a lake in the distance. Suppose further that all the propitious conditions for veridical perception are present: the traveler's vision is not defective, there is adequate light, and so on. Even so, if he has ever been deluded by a mirage, he might doubt his vision. The only way to remove his doubts is to approach the lake, take a dip in it, and drink the water. If it feels cool and his thirst is quenched, he can be sure about the truth of his perceptual belief. The Naiyāyika calls this volition leading to successful behavior "*samvādi pravṛtti*," literally meaning "coherent volition." One's volition is realized in action and one gets one's desired object. Thus, there is coherence between the objects of volition and perception. One's perceptual belief that there is a lake in the distance is true if and only if there is a lake at a distance. However, to establish this truth, one needs further corroboration. One's perceptual belief about the lake must cohere with one's other beliefs about water, such as that it quenches thirst, wets a thing, and douses fire, which motivates one to act positively or negatively in a particular situation. This model of ratification is similar to that of a crossword puzzle,[6] which combines moderate foundationalism with coherentism. A belief, to be true, must have a content-to-world fit. Yet, until such a belief, produced by an accredited means, matches other beliefs in the existing network, the belief cannot be known/believed to be true, nor can the means of generation be warranted. The same applies to scientific knowledge. Naiyāyika-s specifically mention Āyurveda, the science of medicine. Āyurveda is considered a science because Ayurvedic prescriptions lead to successful action. When a sick person is cured by following the prescription of her doctor, she infers the truth of the utterance of her doctor and gradually gains confidence in Āyurveda as a science. Scriptural injunctions, too, are justified by following the same model. Scriptural prescriptions, says Vācaspati, are as acceptable as medicinal prescriptions, since both are uttered by an infallible speaker. It is interesting to note in this context that one of the premises in the argument that God is infallible ultimately rests on another accredited belief-generating process, viz., reliable testimony or authority. Again, when the authority as a means of valid cognition is questioned, there is a fallback on inference. Naiyāyika-s pay scant attention to

6 The clues of the puzzle are like experiential evidence, and other already filled-in entries are like supporting reasons. A particular entry must fit its clue, but it must also match the other filled-in entries. Similarly, in the case of one's belief that *p*, one needs the support of experiential evidence in favor of *p*, and at the same time one needs to see how well this belief *p* fits in one's web of beliefs. To give this type of justification, one need not ascend to an exalted point; nor does one need any *a priori* warrant to establish the adequacy of this justification process (Haack 1990).

obviously circular reasoning. This cannot be due to ignorance. Rather, like true naturalists, they favor repairing their boat while still floating. And, significantly, at no stage in their rebuttal of skepticism do they resort to an internalist mode of justification.

Mādhyamika Buddhists (Nāgārjuna 1995) point at a more basic circularity in the Nyāya epistemology. Nāgārjuna questions the method by which a Naiyāyika attempts to show the reliability of an accredited source of veridical cognition (*pramāṇa*). One cannot establish its reliability and authenticity by means of another *pramāṇa*, for that will lead to an infinite regress. Nor can one establish it by pointing out its reliable character, because that will lead to circularity (an attempt to establish the reliability of a *pramāṇa* by referring to the fact that it has always been sufficient for generating a *pramā* is obviously circular). This objection would be damaging if Naiyāyika-s had only one kind of *pramā* and one kind of *pramāṇa* in their epistemic repertoire. But as Naiyāyika-s admit four different *pramāṇas* for four different types of *pramā*, they can always resort to other *pramāṇas* when the reliability of one is questioned: to justify perception, for example, one might take recourse to inference; and, again, to justify inference, one can rely on verbal cognition; and so on. Quine once said that such scruples against circularity have little point once we have stopped dreaming of deducing sciences from observation. A naturalist need not be afraid of circularity.

Moral Naturalism

Moral naturalism holds that (1) moral values can be understood, if not defined, in terms of moral facts and (2) the borders of the moral and the natural are porous and therefore the fact/value dichotomy is not sacrosanct. Hence, facts of one realm can causally influence facts of the other. Indian understanding of morality falls within a broadly naturalist framework, since both of these features are present in many Indian philosophical systems.

Let us first look at the values admitted in the orthodox systems of Indian philosophy. These are the four main goals of human life, the *puruṣārtha*-s (Banerjee 2008), as, viz., *artha*, *kāma*, *dharma*, *and mokṣa*. The first two are naturally pursued by humans, but the other two are to be pursued because this is enjoined by the scriptures. However, nowhere in the account of these ends is the is/ought distinction pronounced. Rather, there is a smooth transition from what is done to what should be done.

"*Artha*" generally means wealth. However, in the *Kāmasūtra* (2.9), "*artha*" has been defined more broadly as the acquisition of learning, land, gold, cattle, rice, articles for domestic use, and friends, as well as the increase of what has been acquired. The goals of economics, politics, and civics all fall under the purview of *artha*. The desire for these ends need not be selfish, but may be driven by some altruistic motive.

In the wider sense, the word "*kāma*" means desire in general; in the narrow sense, it means sensual desire or passion. Though sheer lust and carnal craving have been condemned in the religious and philosophical texts, controlled lust is looked upon as a necessary means for the survival of the human race. "*Kāma* begins with desire, in a search for pleasure – but ends in love where the desire is transformed into self-surrender and becomes apprehension of a *rasa* as in aesthetic experience. If *artha* finds institutional stability in a country's political economy, *kāma* finds its expression in art and poetry" (Mohanty 2001, 96).

"*Dharma*," etymologically, means that which sustains the people, or a society. *Dharma* has been understood as acting according to the rules laid down in the scriptures and incorporated in social conventions. The first two ends are to be pursued following these social

and scriptural norms. The practice of *dharma* can check the desires and passions and is conducive to the purification of mind, cultivation of desirelessness, and a sense of equality (*samatva-buddhi*).

The first three goals lay the foundations of a well-coordinated social life, but human excellence lies in pursuing liberation (*mokṣa*), which takes us beyond the conventional realm of dos and don'ts. Different philosophical systems have backed different ideals as *mokṣa*. To some, it is freedom from suffering; to others, it is the transformation of one's mundane being into an everlasting blissful being. Some hold it a post-mortem stage; others admit the possibility of attaining it in this life. Those who attain liberation in this life spontaneously perform what is good for all. Whatever way liberation might be viewed, one needs to transcend the cycle of life and death. This can be attained, according to almost all philosophical schools, by eliminating the wrong concepts about the nature of one's self and by acquiring the right knowledge about the nature of reality. Gautama, for example, laid down an exact causal sequence leading to liberation. He writes, "pain, birth, activity, faults and misapprehension – on the successive annihilation of these in the reverse order, there follows release" (1967). To explain, misapprehension, faults, activity, birth, and pain are responsible for our worldly experience. Liberation, which means getting rid of the worldly experience that is sure to lead to suffering, can be attained by the true knowledge of the reality. When the liberating knowledge dawns, one's faults, viz., affections and aversions, disappear; one is no longer subject to any activity and is consequently released from transmigration and suffering.

Some Indian systems, such as Mīmāmsā, endorse moral non-naturalism by admitting two distinct causal realms: ritual and natural. Ritual laws, according to this system, are independent of the natural order and help give an alternative explanation of the causal connection that supposedly obtains between correct performance of a ritual and its result; for example, one desirous of a son should perform a special sacrifice called *putreṣṭi*. However, most philosophers believed in a principle of cosmic order or harmony, viz., *Ṛta*, which "makes science possible, the world beautiful and the humans moral" (Chakraborti 2006). *Ṛta* is the principle underlying the "finely-tuned universe," the transgression of which leads to sin. It represents the totality of physical and moral laws, which even the gods are bound to obey. The law of *karma* follows from *Ṛta* as the causal basis of the phenomenal world. According to the doctrine of *karma*, every action gives rise to some consequence; a good act leads to a good consequence and a bad act to a bad consequence. Every human agent reaps the consequences of his/her actions in this life or in some future life. Not only that, but a just moral scheme requires that one should never suffer/enjoy the consequences of another's actions. Thus is the doctrine of *karma* closely associated with individual morality. The burden of moral responsibility for one's deeds is borne by the individual, fair and square; no one escapes the law of *karma*. In spite of this, most Indians believed and still believe that even if one's present actions are causally necessitated by one's past actions, one's present actions can be free and the future self-determined. Commitment to compatibilism, once again, is a mark of moral naturalism.

Naturalism of the Nyāya-Vaiśeṣika system appears to be compromised by the admission of the unseen force or *adṛṣa*; the methodological naturalism of different philosophical systems is also weakened if we cannot explain how *mokṣa* or liberation can be the ultimate end of both science and philosophy. Since Indian philosophers have admitted an unbroken chain of causal connection within the empirical world spanning across different lives, the unseen force also falls well within the natural world, as is obvious from our discussion of the

cosmic principle of harmony. God, too, is constrained to create the world in consonance with the accumulated *karma* of individuals. Besides, this overarching cosmic principle warrants that, ultimately, the natural and moral orders coincide. Metaphysics, science, and morality thus forming part of the same continuum strengthens naturalism in most Indian philosophical systems.

References

Banerjee, S.C. (2008). Purusartha. In *Philosophical Concepts Relevant to Sciences in Indian Tradition*, vol. III, pt. 5, edited by P.K. Sen. New Delhi: PHISPC.

Barrett, R.B. and Gibson, R.F. (1990). *Perspectives on Quine*. Oxford: Basil Blackwell.

Bhattacharya, K.C. (1956). Studies in Samkhya Philosophy. In *Studies in Philosophy*, vol. I, edited by G. Bhattacharya. Calcutta: Progressive Publishers.

Bhattacharya, R. (2002). Jayantabhatta's Representation of the Cārvāka: A Critique. In *Indian Culture: Continuity and Discontinuity: In Memory of Walter Ruben (1899–1982)*, edited by Joachim Heidrich, Hiltrud Rüstau, and Diethelm Weidemann, pp. 85–93. Berlin: Trafo.

Chakraborti, S.C. (2006). Rta, Satya, Tattva and Tathya. In *Philosophical Concepts Relevant to Sciences in Indian Tradition*, vol. III, pt. 4, edited by P.K. Sen. New Delhi: PHISPC.

Chatterjee, A. (2006). Kārya-Karaṇa-Bhāva. In *Philosophical Concepts Relevant to Sciences in Indian Tradition*, vol. III, pt. 4, edited by P.K. Sen. New Delhi: PHISPC.

Chatterjee, A. (2012). Naturalism in Classical Indian Philosophy. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/naturalism/ (last accessed July 15, 2015).

Chisholm, R.M. (1992). *Theory of Knowledge*, 3rd edn. New Delhi: Prentice Hall of India Private Limited.

Danto, A.C. (1967). Naturalism. In *Encyclopedia of Philosophy*, vol. 5. New York: Macmillan and The Free Press.

Dasgupta, S.N. (1987). *Natural Science of the Ancient Hindus.* Delhi: Indian Council of Philosophical Research and Motilal Banarasidass.

Dharmarajaddhvarindra. (1970). *Vedānta Paribhāṣā*, edited and translated by Pancanana Bhattacharyya Sastri. Calcutta: Sanskrit Pustak Bhandar.

Ganeri, J. (2012). *The Self: Naturalism, Consciousness, and the First-Person Stance.* Oxford: Oxford University Press.

Gangeśācārya. (1974). *Tattvacintāmaṇi, Pratyakṣa-Khaṇḍa*, edited by Kamakhyanatha Tarkavagisa. Delhi: Motilal Banarasidass.

Gangopadhyay, M.K. (2008). Paramāṇukāraṇavāda. In *Philosophical Concepts Relevant to Sciences in Indian Tradition*, vol. III, pt. 5, edited by P.K. Sen. New Delhi: PHISPC.

Gautama. (1967). *Nyāyasūtra* (with Vātsyāyana's *Nyāyabhāṣya,* Uddyotakara's *Nyāyavārttika,* Vācaspati's *Tātparyatīkā,* Udayana's *Pariśuddhi*), edited by A. Thakur. Darbhanga: Mithila Institute.

Haack, S. (1990). Rebuilding the Ship while Sailing on the Water. In *Perspectives on Quine*, edited by R.B. Barrett and R.F. Gibson, Oxford: Basil Blackwell.

Hepburn, R.W. (1967). Nature, Philosophical Ideas of. In *Encyclopedia of Philosophy*, vol. 5, edited by Paul Edwards. New York: Macmillan and The Free Press.

Iṣwaṛakṛṣna. (1972). *Sāṃkhya-kārikā,* with the commentary of Gauḍapāda, translated by T.G. Manikar. Poona: Oriental Book Agency.

Kitcher, P. (1983). *The Nature of Mathematical Knowledge.* New York: Oxford University Press.

Kornblith, H. (1985). *Naturalizing Epistemology.* Cambridge, MA: MIT Press.

Mādhavācharya. (1882). *Sarva-darśana-saṃgraha*, edited by E.B. Cowell and A.E. Gough. London: Trubner.

Matilal, B.K. (1986). *Perception: An Essay on Classical Indian Theories of Knowledge.* Oxford: Oxford University Press.

Mohanty, J.N. (1992). *Reason and Tradition in Indian Thought.* Oxford: Clarendon Press.

Mohanty, J.N. (2001). *Explorations in Philosophy: Indian Philosophy.* New Delhi: Oxford University Press.

Nāgārjuna. (1995). *Mūlamadhamakakārikā* (The Middle Stanzas). In *The Fundamental Wisdom of Middle Way*, edited and translated by J.L. Garfield. New York: Oxford University Press.

Papineau, D. (2009). Naturalism. In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Available from: http://plato.stanford.edu/entries/naturalism/ (last accessed July 15, 2015).

Praśastapāda. (1930). *Padārthadharmasaṃgraha with Commentaries, Śukti, Setu and Vyomavatī*, edited by G. Kaviraj. Varanasi: Chowkhamba; and (1963). With *Nyāyakandalī* of Śrīdhara, edited by D. Jha. Varanasi: Sanskrit Viśvavidyālaya.

Quine, W.V. (1969). *Ontological Relativity and Other Essays.* New York: Columbia University Press.

Seal, B. (1958). *The Positive Sciences of the Ancient Hindus.* Delhi: Motilal Banarasidass.

Strawson, P.F. (1985). *Skepticism and Naturalism: Some Varieties.* London: Methuen.

Udayana. (1996) *Ñyāyakusumānjali*, edited and translated by N.S. Dravid. Delhi: Indian Council of Philosophical Research.

Umāsvāti. (2007). *Tattvārthādhigamasūtra*, translated by N. Tatia. Delhi: Motilal Banarasidass.

Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*, translated by D.F. Pears and B.F. McGuinness. London: Routledge and Kegan Paul.

35

The Natural History of Shame and its Modification by Confucian Culture

RYAN NICHOLS

The purpose of this chapter is to offer a natural history of shame. Philosophical discussions of nature and naturalism are befuddled by false dichotomies, such as nature versus nurture, or the natural versus the cultural. Since culture is part of nature, we examine the utility of a sense of shame and shame emotions from the point of view of the evolutionary history of social primates; I continue the story by discussing the social and cultural transformation of shame emotions through Confucian Chinese culture. This chapter shows why the dichotomy between nature and culture is a false one. More important, the chapter argues that several interacting forces, from physical ecology to cultural transmission of written texts, are key events in the natural history of shame.

Shame as a Behavioral, Emotional, and Cognitive adaptive System

What is shame? This section examines what shame experience is through consultation of relevant studies in evolutionary and social psychology, and finds it is a "social rank-based emotion." It then examines why shame is thought to be an adaptation not only in our species but in our superfamily *hominoidea*.

What Is Shame?

The English term "shame" can connote a *sense* of shame, *feelings* of shame, shame *behavior*, and the *state* of being ashamed. More clearly, shame is an "affect," and for some a basic emotion (Gilbert 1997). Unlike other emotions, affects are considered to be cross-cultural natural kinds (Griffiths 1997, 4), following what Ekman (1992) says about "basic emotions."

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

Benefiting from the wide interdisciplinary literature on shame and guilt requires resolving conceptual confusions regarding the senses and referents of those two terms. Jeff Elison signals the problem by titling a paper "Shame and Guilt: A Hundred Years of Apples and Oranges" (2005). Work clarifying the categories involved reveals an insight constraining the present project: conceptions of shame in psychology and philosophy are far too complex either to represent experimental data about shame or to correspond to the evolutionary ancestry of shame (Elison 2005, 6).

Studies document correlations between shame and the basic emotion of fear. Data drawn from an fMRI study reveal similar activations in the anterior cingulated cortex during both physical pain and social exclusion (Eisenberger, Lieberman, and Williams 2003). Shame's capacity to change behavior proceeds through fear or the cognizance of fear, but depends upon social intelligence.

Experts on emotion concur: shame is primitive and requires no cognitive resources for its experience. This contrasts with guilt. Fessler describes shame as having "evolved from a rank-related emotion," which explains why it can be found "motivating prestige competition, cooperation, and conformity" (2004, 207). Gilbert writes, "The evolutionary root of shame is in a self-focused, social threat system related to competitive behavior and the need to prove oneself acceptable/desirable to others" (2003, 1205). He concludes that it is the "inner experience of self as an unattractive social agent, under pressure to limit possible damage to self via escape or appeasement, that captures shame most closely…Shame is an involuntary response to an awareness that one has lost status and is devalued" (1998, 22). Shame experiences monitor social rank and detect social threats from conspecifics. Decoding the social communication as disapproval enables the shamed animal to submit. Sometimes, this decoding is fully conscious. Sometimes, the shame system activates unawares, as in blushing: you might experience shame and then blush without being aware you are blushing, enabling *others* to know you have been shamed even when *you* don't.

The capacity for shame evolved in hominins and non-hominins due to its utility in enabling those who have it to negotiate dangerous social hierarchies (Gilbert and McGuire 1998). Shame is elicited by one's perception of one's social devaluation by another (Frijda 1993; Gilbert 1997). Thus, shame experience represents a "recalibrational emotional program," in this respect like guilt, grief, depression, and gratitude. The primary function of recalibrational emotions is not short-term changes to behavior (Tooby and Cosmides 2008).

In human cultures without enforced laws (such as Early China) and in various non-human contexts, shame or "protoshame" is one of few nonviolent means by which the group can regulate behavior (Fessler 1999). Shame in non-human primates takes forms such as averted gaze, reduced posture, and submissive gait, stereotypic features of appeasement. Fessler remarks that such displays "signal to dominant individuals that the actor accepts a subordinate position in the dominance hierarchy"; these gestures aim to "dissuade the dominant from aggressing" (2004, 239; see Fessler 1999; Keltner and Harker 1998). Fessler adds that the "panprimate substrate" on which human shame is built consists of communication within a system of hierarchical social relationships; this, he argues, changed in human beings because the reliance on culture and cooperation favored the evolution of a new motivational system oriented toward prospective cooperative partners (2007, 174). Shame's role in catalyzing appeasement displays is a homology shared between present-day humans and our hominin ancestors. Humans unable to appease dominant individuals after violations of group norms risk being targeted by collective action. Keltner and colleagues find that displays of shame enhance reconciliation and social reincorporation (Keltner and

Harker 1998). Conformity is motivated by a desire to reconcile or to avoid being ashamed and embarrassed (see Fessler 2004). Communicating that the actor knows and values the local standards mitigates the damage the violation causes to the actor's social position (Keltner and Buswell 1997).

Why Is a Sense of Shame Thought to Be an Adaptation in Social Primates?

Where social leaders use physical aggression to enforce rank and harmonize behavior, such leaders are coercive, threatening, and authoritarian (Keltner and Harker 1998). But where prestige is the currency through which social leaders enforce rank and harmonize behavior, such leaders emphasize the desirability of talent, competence, and affiliation. Leaders of prestige-based social groups, "attractors," lead not by instilling fear and obedience, but by being valued and popular. Gilbert and McGuire (1998) refer to this as "social attention-holding power." If prestige is the currency of social rank, then the physical correlates of shame behaviors will change by becoming less severe and pronounced, while retaining the structure imposed on them by the utility of shame displays throughout our species' biological evolution.

Whether prestige or physical dominance is the means of ordering society, society will remain hierarchical. Shame experiences function to appease dominant individuals in the hierarchy and to signal submissiveness (Keltner and Buswell 1997). In humans, ultrasocial creatures that we are, shame displays were redirected in cultural forms from aggression-avoidance to the maintenance of prestige. Our species expanded the range of behavioral contexts in which shame became appropriate (Gilbert 2003, 1209). Prestige is a proxy for possession of access to resources. Prestigious people are models of social success and influence, which causes lower-ranking individuals to concern themselves with and to imitate prestigious people (Henrich and Gil-White 2001). Decreases in prestige are maladaptive. Shame displays limit or halt such decreases.

Shame displays also enforce conformity. Fessler remarks about what was earlier referred to as culture's role in changing the bioprogram that "Natural selection therefore modified a fundamentally competitive emotion so that it could also motivate simple conformity – in addition to striving not to be bested by their rivals, individuals now also worked not to fall out of line with the majority" (Fessler 2004, 250). Conformity to group standards not only signals group membership, which increases access to resources, but also enables individuals to "more readily predict one another's actions. Increased predictability is valuable because it facilitates coordination across individuals" (2004, 245).

Since status and rank are positively correlated with adaptive reproductive benefits (Fournier, Moskowitz, and Zuroff 2002), signaling submission to an authority, a group, or a parent is highly adaptive (Gilbert and McGuire 1998). Stereotypical voluntary or semi-voluntary signals of submission include making oneself appear nonthreatening, surrendering, and physically lowering oneself. Participants across cultures successfully identify shame behavior as such, closing a gap that would lead to misunderstanding if it were left open. In many social primates, shame assists in the management of the hierarchy without negative consequences from physical aggression on groups (Keltner and Harker 1998). Submissive signaling represents one's recognition of social devaluation by others and represents a warning signal to oneself of one's social devaluation (Gilbert and McGuire 1998). Data on the phenomenal components of shame across human cultures show that shame consists of feeling weak, lower, and small (Barrett 1995).

Causing people to feel inferior in these and other ways can be of use for leaders, since these feelings can be converted into behaviors. Shame and other self-conscious emotions, as Parker puts it, "were favored because they increased the ability of parents, kin, peers, and others to socially manipulate their offspring, kin, peers, and others. Self-regulation functions of [the self-conscious emotions] were favored because they allowed parents and others to inculcate and enforce values and behavioral codes in others that increased their own inclusive fitness" (Parker 1998, 115). This "social manipulation" can be studied through the lens of the mechanisms that facilitate the transmission of cultural information from one generation to the next. Various groups increase the group's net reproductive fitness relative to other groups through conformist transmission. This is because high rates of conformity minimize differences within the group and aid cooperation, while widening differences between groups (Boyd and Richerson 1988). Conformist transmission will be most effective in physical and social ecologies that are stable, and will be least effective in environments that are unstable (Henrich and Boyd 1998, 220, 226). For numerous reasons, the ecology of Early China was a remarkable incubator for conformist transmission via shame.

Shame in Early Confucianism

The previous section analyzed shame in the context of its natural and evolutionary history in social primates. The central finding was that shame construed as a social rank-based emotion was selected for because of its fitness-enhancing or fitness-preserving qualities. When shame makes the jump from non-human social primates to acculturated human beings, the natural history of shame enters a new chapter.

By making a few assumptions, we treat Confucian culture as part of nature. First, if metaphysical naturalism is true, everything that exists is in spacetime. This includes culture. Since culture is part of nature, cultural change can be studied scientifically. In the present context, this prompts us to better understand the means by which human culture has altered our inherited biology and psychology. Even among those who endorse metaphysical naturalism, the present point is regarded as much more controversial in the biological and evolutionary sciences than it is, perhaps, in the humanities. Yet scientists document key changes wrought on our species – even on its genome – by nothing other than culture. Consider that a subset of humans have a gene facilitating lactose absorption. Studies of the natural history of this single gene show that the various points in our species' natural history at which it arose were all correlated with cultural practices (Simoons 1969; 1970). This gene was not selected unless herding behavior, a cultural practice requiring information transfer between human beings, was present. Cases such as this inspire a naturalistic approach to culture's interaction with features of our suite of biological, psychological, and social affordances: our "bioprogram," for short.

Suppose the foregoing assumption is correct and culture is part of nature; is there a scientific means to study culture? This section's assumption is that cultural transmission theory is currently the best theoretical framework within which to study cultural change (Boyd and Richerson 1988; Henrich and Boyd 1998). Cultural transmission theory asserts that "culture" is best understood as information stored in brains and conveyed horizontally, at one time, or vertically, across generations. This is resonant with the extended mind hypothesis (Clark and Chalmers 1998). According to this theory, cognitive processes are not all within the head. Cultural transmission theory makes use of this point. Since artifacts can be

fashioned into repositories for the storage of information by humans and non-humans, transmission of the fruits of cognitive processes serves to efficiently transmit information vertically across generations. Cultural transmission theory hypothesizes that several key mechanisms involved in the transmission of this information can be tracked in accord with social scientific laws (see Boyd and Richerson 1988). In the present context, this raises the prospect of developing a scientific understanding of the natural history of shame as a social rank-based emotion in social primates, in conjunction with an understanding of a specific culture's use or modification of that natural affordance.

It is to this we now turn in order to describe the contexts and contents of Early Confucian culture relevant for understanding how Confucian cultural practices modified the human shame system. In this section, we first describe the violence and social disarray in Warring States China, and motivate the inference that Confucians leveraged shame to change that state of affairs. We then support this inference with evidence of Early Confucians' preoccupation with prestige and social rank. Finally, we look at Early Confucianism's use of this shame to increase quasi-moral behavior and cooperation.

Social Disarray and Violence in the Warring States

Where rank is determined by aggression, physical prowess is a key virtue (Latin *vir* = man; manly). Where rank is determined by social attention-holding power, social influence and empathy are prized. Early Confucian texts aimed to increase social harmony and decrease violence by promulgating a networked ideology with specific emotional, behavioral, and cognitive content. Confucians in Warring States China inherited a social rank system dominated by physical aggression. In response, they first recalibrated the sense of shame to facilitate development of a social environment in which dominance is achieved by prestige, not force. Second, they altered shame experience and the sense of shame so that embodiment of a suite of quasi-moral traits of character was necessary for achievement of high social status. In the sense of the term "prestige" used by cultural transmission modelers Joe Henrich and Francisco Gil-White (2001), Early Confucians aimed to create a system whereby prestige was earned through morally praiseworthy behavior. To understand why this came about, we must first know something of the social and physical ecology of Early China.

Due to a few historical and ecological features of China, as well as the settlement of the North China Plain, subsequent populations experienced high rates of violence and improvable rates of cooperation. The two most important of these features were the early development of agriculture (in the 7th millennium B.C.E.) and the history of metallurgy in East Asia. The advent of agriculture changes population dynamics and social organization, prompting the development of hierarchies. China's early adoption of agriculture speeded and amplified its social stratification. As agriculture allows societies to gain increasing rates of organization, conflicts of interest between individuals are settled by the powerful exploiting the weak to maximize fitness at others' expense. Laura Betzig finds significant correlation between group size and hierarchical complexity, asymmetry in the resolution of conflict, and rate of polygyny attained by those at the top of the hierarchy (1982; see Betzig 1995).

The second major incubator for social violence in Early China was the onset of metallurgy. The importation of metallurgical skills furthered Early China's stratification (Higham 1988). Kyle Summers notes unique features of the Early Chinese context that led to "extreme hierarchies" (2005). While the Shang Dynasty government had a monopoly on bronze metallurgy, peasants were left to use stone tools for agricultural work. Agriculture allows for

the accumulation of wealth, which in turn separates levels of society. Access to metal implements for farming and metal weapons engenders further separation between levels of society, since the highest levels – those with access to metal implements and weapons – vastly outcompete levels without access.

These facts play indirectly into the story of shame in Early China and Confucianism for two reasons. First, societies highly stratified by wealth fall victim to more social unrest and violence than societies with more egalitarian forms of resource distribution. Second, extensively hierarchical cultures make more use of shame and deference than do less hierarchically organized cultures (Kitayama, Markus, and Kurokawa 2000). "In more interdependent and hierarchically structured societies there is often a hypercognized emotion, combining elements of what we call shame, embarrassment, shyness, and modesty, which is highly valued when displayed by the lower-status person in an interaction" (Keltner et al. 2006, 131; see Menon and Shweder 1994).

The Warring States period (*zhanguo shidai*, 戰國時代) earned its name. Historians, sociologists, and anthropologists of Early China studying the periods after the fall of the Western Zhou Dynasty in 771 B.C.E. to Qin Unification in 221 B.C.E. emphasize the extreme rates of interpersonal and intergroup violent conflict, much of which traces to shame through dishonor. The *Zuozhuan* lists 540 interstate wars and 130 major civil wars in the North China plain across 259 years (Lewis 1989, 36). The aristocracy lived in a nearly constant state of war. Since warfare represented the defense of the clan's lineage and honor, it became a part of the ancestral cult, institutionalized as a religious duty. Bellah argues China's "axial transition" occurred "when a society ruled by warriors was being transformed into a society ruled by imperial bureaucrats"; this occurred during "the Warring States period (450–211 B.C.E.) when a series of new developments changed the nature of Chinese culture and society and led to the elimination of the warrior aristocracy that still dominated in the Spring and Autumn period" (2011, 401). The central cultural development Bellah goes on to describe is the advent of Confucianism. Experts in Early China concur (Pines 2002, 89). During the transition from the Spring and Autumn period to the Warring States period, Confucius is born and lives, marking the beginning of a broad cultural change.

Concern for the peaceful ordering of Chinese society while maintaining the social hierarchy informs some of the earliest layers of Confucius' *Analects*, a text built up in many accreted layers, and some of the earliest dialogues recorded in *Mencius* (c. 300 B.C.E.). *Mencius* Book 1A.6, for example, reports the death of King Xuan and the ascension to the throne of his son Xiang, then describes Mengzi's first meeting with the newly appointed sovereign. The king asks, "How can the world be pacified?" The dialogue continues:

> I responded, "It can be pacified by being unified."
> The king asked, "Who can unify it?"
> I replied, "One who does not have a taste for killing people can unify it…If there were one who did not have a taste for killing people, the people of the world would crane their necks to look for him." (*Mencius* 1A.6; quotations from *Mencius* are from Van Norden 2008, unless noted)

This overriding concern arises in the opening section of the second most important text in the Confucian tradition, carrying immense cultural and historical influence through subsequent generations. At *Mencius* 4A.14 we read, "Kongzi rejected those who enrich the rulers who do not put into effect benevolent government. How much more would he reject those who encourage war. To wage war to fight for land is to kill people till they fill

the fields. To wage war to fight for cities is to kill people till they fill the cities. This is what is meant by 'leading the land to eat the flesh of the people.' This is a crime even death cannot atone for." Leaders who promulgate war and violence are the worst of the worst. Through occupancy of key advisory roles in the civil service, Confucians sought to shift the causes of social governance from aggression and force to create a more peaceful, cooperative society.

Achieving Social Status: The Goal of a Good Confucian

According to the foregoing argument, Early Confucianism bases prestige not on physical power and capacity for violence but on quasi-moral character. This implies that Early Confucians aim to achieve social status. Early Confucian texts reveal an extreme concern for achieving social status, often through the public display of virtue. Confucius affirms the publicity of virtue explicitly: The Master said, "Virtue is never solitary; it always has neighbors" (*Analects* 4.25; quotations from the *Analects* are from Slingerland 2003, unless noted). Confucianism was able to use shame to enforce or encode behaviors of would-be scholar officials (*shi*, 士), the target audience of Early Confucian texts, because the code to which these readers aspired correlated socially desirable behavior with moral behavior.

Consider first *ren* (仁), benevolence, thought the highest virtue. The subtle opening statements in the earliest layer of the *Analects*, ch. 4, exhibit a means–end concern with virtue and goodness while combining an interest in status. *Ren* is a means, though not a mere means, to other things of more importance: fame, overcoming adversity, and being happy.

Consider next the *junzi* (君子). In the Confucian texts, this is typically translated "gentleman" or "superior man," but the term jun (君) originally referred to a man of high social rank, a "prince." Put otherwise, Early Confucianism's moral ideal, the *junzi*, was itself defined in terms of preeminent social status (one obtained by moral accomplishment, not by birthright). Since the *junzi* is characterized as someone of high social rank, Early Confucianism leverages the *junzi*'s public persona to motivate moral behavior (*Analects* 19.2). *Daxue* (*The Great Learning*, 大學) says *junzi* ought to be wary of being alone and not in public (Chan 1969, 89–90), as does *Zhongyong* (*The Golden Mean*, 中庸; Muller 1991/2013, 1). Presumably this is because in a shame-based moral system, being unsupervised undermines the primary form of moral motivation at work.

Historian Ping-Ti Ho remarks that Early Confucians transformed their concern with increasing their social status in a highly stratified society into moral distinctions between kinds of men. (Note that due to the content of these texts, use of gender-inclusive language to discuss the social structure of ancient China would misrepresent the contents of that structure.) This led Early Confucians to alter the meaning *junzi*, which they shifted from referring to the nobleborn to a person "of superior knowledge and moral quality" (Ho 1962/2008, 5). These authors and editors sought to maintain social hierarchy but shift its principle of organization away from power (passed on through kin relations) and toward moral character.

Early Confucian cultural leaders used rank-based emotions as a tool for a sweeping set of social policies that would overcome the endemic violence in the Spring and Autumn and Warring States periods. Seeding a sense of shame in the population of scholar officials, and, through them, in their students, employees, and subjects, and, through *them*, in their sisters, brothers, sons, and daughters, had greater potential to peacefully achieve social harmony than just about any other method of social engineering available.

Early Confucian Texts Use Shame to Promote Nonviolence and Cooperation

Philosophers Rosemont and Ames say that shame was "a hugely important factor in Confucian moral philosophy," since it enabled transformation of "the people through their own willing participation in a self-regulating communal order [which] is fundamental to Confucianism" (2009, 58, 73). This brief subsection pauses to describe a few techniques identified in Early Confucian texts for promoting social harmony through natural mechanisms that involve shame. Early Confucians routinely exploited physiological responses to situations (like blushing) for the benefit of their adherents, or specialized in crafting situations (like rituals) that prompted physiological changes in the body and brain. In both cases, Confucian practice molds our bioprogram to its own ends; specifically, toward social harmony.

Face-Reading and Mind-Reading: Learn to Identify Shame Experience in Others

Fingarette says about shame (*chi*, 恥) that it "looks 'outward,' not 'inward'…The Confucian concept of shame…is oriented to morality as centering in *li*, traditionally ceremonially defined social comportment, rather than to an inner core of one's being, 'the self'" (1972, 30). The outward focus of Confucian shame is possible because Early Confucians appealed to embodied signals of shame and shamelike feeling as a means of success in theory of mind tasks (*Analects* 2.8). In *Mencius* 3B.7, Zilu describes how you can gain valuable information about the feelings of one who utters what he does not believe: "observe the reddening of their facial coloration" (Csikszentmihalyi 2004, 135). Blushing and sweating signal inauthentic behavior, knowledge the disciple of Confucianism can use to his advantage. Savvy Confucians expertly perceive when people reduce their social rank, and train in advanced techniques of impression management.

Propriety Produces Cooperation: Use Rituals to Bring People together Peacefully

Early Confucianism converts the complex of emotions, behaviors, and cognitions surrounding role-based shame into actual physical behavior by prescribing a set of preferred ritual interactions, and representing negative consequences of violations of those prescriptions. Ritual propriety (*li*, 禮) efficiently encodes a sense of shame. Its benefits are obvious: following *li*, you would know better than to greet a visitor in hunting attire, or to fail to serve him turtle soup. These are both acts for which noblemen were murdered in Warring States China (Lewis 1989, 40). "Being respectful brings one close to ritual propriety, and keeps one away from shame and humiliation" (*Analects* 1.13), especially when interacting with other strata in the social hierarchy. Savvy Confucians use rituals that signal deference as a means of social governance at all strata of their well-ordered society.

Shame Is the Preferred Means of Social Control: When One Must Actively Govern, Govern without Force

Consistent with *Analects* 2.11 and *Mencius* 2A3.2, pre-Qin and Imperial Chinese states used forms of nonviolent punishment (in addition to other forms) that relied on shame and ostracism, including tattoos and public humiliations (Ch'ü 1961, 246–247). These punishments drew from both Early Confucianism and a cotemporaneous, competing philosophical school, legalism. The Qin Dynasty – influenced considerably more by legalists – contained a legal canon providing conditions for appropriate tattooing, which leveraged fear and

conformity to enforce norms (McLeod and Yates 1981, 147). Since the body was inherited from one's ancestors, its disfigurement through amputatory punishment caused family and euhemerized ancestors to feel shame (Rosemont and Ames 2009, 12). Qin and Han law provided for "mutual implication" (*lianzuo*, 連坐), according to which punishment for some major crimes didn't just focus on the violator but "extended to family, neighbors, and, in the case of an official, to superiors, subordinates, or the man who had recommended him for office" (Lewis 2010, 233; see Cook 2004, 407). Consequently, the costs of third-party punishment and policing were reduced dramatically. Savvy Confucians create social harmony with minimal physical punishments for offenders by deft use of shame's social utility.

Early Confucianism's Influence Present in a Unique Shame Profile in the Present-Day Confucian Diaspora

If the Confucian cultural tradition has the content this argument says it does, and if it has been as influential and well-transmitted as this argument presumes it has, then we would expect that people in the present-day Confucian diaspora will show evidence of heightened shame behavior, shame emotion, and shame cognition in relation to control groups not subject to the ideology of Early Confucianism. The first part of this section shows that members of the Confucian diaspora feel heightened shame experience compared with WEIRD and other non-WEIRD countries (Western, Educated, Individualist, Rich, and Democratic; see Henrich, Heine, and Norenzayan 2010). The next part shows the ways in which civil and political authorities, parents and pedagogues in the Confucian diaspora have employed the social utility of shame.

Confucian Diaspora Populations Have a Unique Shame Profile

When a culture modifies evolutionary norms having to do with shaming and being shamed, its social communities are led to endorse some behaviors and condemn others. Continuing methodological problems in social psychology arising from its nearly exclusive use of WEIRD subjects prevent us from definitively drawing this conclusion, since single shame studies applying the same design to East Asian subjects, WEIRD subjects, and subjects who are both non-East Asian and non-WEIRD are rare. Nonetheless, many cross-cultural studies of shame show that East Asians on average feel more shame, in more contexts, and more frequently, than others (Sznycer et al. 2012). Chinese have higher needs for socially oriented achievement, abasement, change, endurance, and order, and lower levels of individually oriented achievement, affiliation, aggression, exhibition, and power than do others (Bond 1986). Keeping in mind what shame experience is – a social rank-based emotion and corollary set of behaviors and cognitions – studies of several of these features point to a heightened salience of shame experience in Chinese populations.

Chinese experience more social anxiety than do Westerners (Zhong et al. 2008). In general, East Asians experience "socially engaged" emotions more intensely than Westerners (Kitayama et al. 1997; Kitayama, Mesquita, and Karasawa, 2006). European Americans experience shame predominantly in response to something they themselves do, but East Asians experience shame in response to either their own or a close friend or relative's actions (Wong and Tsai 2007). In interdependent and collectivist cultures, family members, friends, and deceased ancestors are experienced as extensions of the self (Markus and Kitayama

1991). Supporting data come from Stipek (1998), who found that Chinese students were more likely than American students to feel guilt or shame in response to a relative's wrongdoing, as well as pride for the praiseworthy accomplishments of a relative. Stipek also asked Chinese and American participants to rate the degree of shame they would feel if they were caught cheating and the degree of shame they would feel if their brother were caught cheating. Chinese reported they would feel significantly more shame in the event that their brother was caught cheating than did Americans.

Language and narrative in China appear to emphasize shame more frequently than elsewhere. A study of Chinese children's stories rank-ordered the frequency of portrayed behaviors as follows: socially oriented achievement, then altruism, then social and personal responsibility (Blumenthal 1977; cited in Stipek 1998, 619). A more recent study with comparable design found that Chinese storytelling focused on morality and elicited shame, whereas U.S. storytelling focused on entertainment and elicited positive emotion (Miller et al. 1997). Relatedly, Chinese children learn shame terms at an earlier age than do children in England and the United States (Shaver, Wu, and Schwartz 1992). Li, Wang, and Fischer (2004) found 83 terms having to do with shame in Chinese. This provides further confirmation that shame is hypercognized through the Chinese language.

Surveys of cultural values based on self-report measures provide a snapshot of the variance of values across cultures. The value "Having a Sense of Shame" ranked 9th of 40 different values in the Chinese Values Cultural Survey (reported by Garrott 1995, 219). Studying Chinese shame concepts, Li, Wang, and Fischer (2004) found robust relations between fear and avoidance of shameful experiences. The extent of these relations shows China is exceptional even among collectivist cultures. Studies of emotion categorization show that China's emphasis on honor, respect, and face leads to highly articulated scripts for expression of and reaction to shame (Li, Wang, and Fischer 2004). Shame emerges as an emotion family in China, but not in other collectivistic cultures, such as Indonesia or Italy (Shaver, Wu, and Schwartz 1992; see also Shaver, Murdaya, and Fraley 2001). Tsai (Wong and Tsai 2007) found that Hong Kong Chinese participants valued shame more than Asian American and European American participants.

Residents in the Confucian diaspora exhibit a cross-culturally unique shame profile as compared with populations from outside the Confucian diaspora, especially Westerners. This profile consists of analogs of emotional, cognitive, and behavioral features revealed in Early Confucian texts, including greater shame sensitivity, greater linkage of shame with social fears and social anxieties, and greater identification with one's role, title, and group.

Shame's Utility for Social Governance, Political Order, Parenting, and Pedagogy in China

Once a sense of shame is internalized and once one knows the requirements for maintaining good standing in the community, group action gains efficiencies. Shame reduces a group's resource allocation for third-party justice by governmental bodies, including police, courts, and jails, and shame facilitates increased rates of within-group communication and cooperation. However, this comes with a proviso: people must *care* what others think of their behavior, and people must *pay attention* to others' behavior. As Bedford and Hwang write, "shame is a more effective means of social control in a system where maintaining harmony in relationships is valued over maintaining behavior according to an objectively defined right and wrong" (2003, 133). Rather than query the implications of these points

on cross-cultural differences in moral philosophy, we look at how shame is used for social control in East Asia.

Historians and philosophers of Chinese law document cross-cultural differences along these lines regarding the roles of shame between East and West. Since the Anglo-American legal tradition emphasizes personal liberty, objective rights, property, and Mill's "harm principle," shaming is regarded as an ineffective means of behavioral control. But historians argue that the Early Confucian emphasis on shame has been a formative force in the development of contemporary East Asian legal systems. In *Tradition of the Law and Law of the Tradition*, Xin Ren writes, "Social conformity in the Chinese vocabulary is not limited to behavioral conformity with the rule of law but always moralistically identifies with the officially endorsed beliefs of social standards and behavioral norms…[T]he Chinese tradition of so-called 'greatest unity' has always attempted to achieve ultimate uniformity of both mind and act within Chinese society" (1997, 6). This provides background to data surrounding the rule of law in China. Before-the-fact socialization, rather than after-the-fact arrest, primarily explains the low rates of social crime in China (Schurmann 1968). "Social crime" refers to crime affecting a victim visible to the assailant, like assault, as opposed to nonsocial crimes, like violation of property rights through Internet piracy. Reflective of Confucius' dictum of *Analects* 2.11 about the social inefficiencies of law, Chinese legal theorist Xiaoming Chen writes, "Law has traditionally been regarded as, at best, irrelevant and, at worst, disruptive of harmony. Neither written codes nor legal institutions and practices have traditionally been seen as very important by the Chinese themselves" (Chen 2002, 51; see van der Sprenkel 1966/2004).

Primary responsibility for keeping the peace in China falls on community norms. The community plays a significant role in solving crimes and even apprehending criminals in China (Zhang 2004). Perhaps the most fascinating example of social enforcement of law and morality lies in contemporary Chinese netizens' use of what is called *renrou sousuo* (人肉搜索; literally, "human meat search engine"). The community vigilance and social application of shame at work in *renrou sousuo* has longstanding precedent in Chinese history. The most formal of these precedents dates from the Song Dynasty and its organized community-policing system of *Baojia* (保甲). East Asian, especially Chinese, societies appear to have been shaped significantly through the cultural manufacture of cooperation and obedience through practices leveraging shame experience.

A similar set of effects is observed in parenting and pedagogy in China. When examining the use of shame in the influence of children in contemporary cultures, we find that the role of shame in instruction or conditioning varies significantly. Research shows that shame is used frequently in China for the moral instruction of children (Fung 1999). According to one cross-cultural study appraising the role of shame in shaping behavior, contemporary China uses shame more often than does other societies – East Asian or Western (Mauro, Sato, and Tucker 1992, 313).

In an exploratory factor analysis of beliefs about and practices of child-rearing in China, Lieber and colleagues found two factors were closely related: *shame* and *training*. Furthermore, *training* was the only factor with which shame was linked in their model. They wrote, "That Shame is significantly related only to Training suggests a dynamic unique within Chinese parenting beliefs and consistent with the socialization of traditional filial values. 'Training' embodies Chinese parents' sense of responsibility to foster the development of pro-social behavioural characteristics and this effort is accompanied by 'shame' practices intended to heighten children's sensitivity to the emotional and behavioural cues

in Chinese social contexts" (Lieber, Fung, and Leung 2006, 145). When to feel shame, how to feel shame, and when to apply shame to others are subject to unique amounts of social training in East Asia, despite the presence of evolved capacities for shame experience. Such fine perceptual attunement of the sense of shame must be produced through cultural learning.

For example, when Western children report that the appropriate situational response to social exclusion is anger, Chinese children report that the appropriate emotional response is shame. These two sets of responses are funneled through social conventions that govern the display of positive and negative emotions in children (Cole, Bruschi, and Tamang 2002). The explicitly social conditioning of shame in Confucian populations is reflected in cultural tropes, including Confucian culture's "emphasis on face, criticism, and evaluation in interpersonal relationships…and in its rich variety of lexical terms and labels for shame, humiliation, and embarrassment, face, and related notions" (Fung 1999, 183). Developing a sense of shame allows Chinese children to be sensitive to the perceptions of others, and thereby to avoid future behaviors that would bring shame or embarrassment to the family. Children learn that shame functions like a perceptual faculty that takes as its object social status vis-à-vis the group, which allows a culture's expression of shame experience to manage social roles (Li, Wang, and Fischer 2004, 769). Nelson infers that "Confucian philosophy is thought to help children regulate and enact their behavior in culturally appropriate, modest, tactful, restrained, respectful, and sensitive ways" (2006, 262). In studies of sets of Chinese and American mothers, observers recorded mothers talking about their child's misbehavior in the child's presence. Chinese parents readily endorsed shaming as a strategy by which to educate and socialize their children about the proper ways to behave (Fung, Lieber, and Leung 2003).

Many, many more features of East Asian societies point to their unique shame profile. For example, shame appears to mediate the diffused sense of self found in East Asia but not elsewhere (Triandis, McCusker, and Hui 1990; Heine 2001; Cohen and Gunz 2002). But these features cannot be taken up here. Confucian diaspora children receive more cultural training from their parents in the use of the cognitive, behavioral, and affective system of shame than do other children. This is expected, given the content and transmission of Early Confucianism.

Conclusion

Confucianism was created by design as a cultural system containing behavioral scripts and rituals to guide interpersonal interaction and emotional schemas that encode responses to social stimuli (Nichols 2011; 2013). At the cognitive level, Confucianism produced a set of texts that were not only faithfully read for thousands of years but were memorized by hundreds of thousands of aspiring men whose reputations – and reproductive success – were de facto dependent on their success in that task via workings of the pre-Imperial and Imperial examination systems. Confucianism began its journey through time by moving an Early Chinese society prone to disarray and violence toward peace and stability. Many widely transmitted Confucian techniques for influence were deployed to enhance our evolutionarily endowed shame system to contribute to a harmonious society. Not only do we know what techniques these were – rituals, proper use of titles, social forms of punishment, and so on – but we have evidence from cross-cultural psychology of their special influence

in the Confucian diaspora today. Reverse-engineering a cultural subcomponent such as Confucian shame has much to teach those of us working in the multidisciplinary world of cultural transmission, who remain discontent with *a priori* mathematical models and laboratory experiments. It has even more to teach those of us who want to understand China from the inside out.

Acknowledgements

Research for this project was provided through a postdoctoral fellowship with the SSHRC-funded Cultural Evolution of Religion Research Consortium (CERC), based at the Centre of Human Evolution, Cognition and Culture at University of British Columbia, and by the Templeton World Charities Foundation, through a sub-award from The Chinese Challenge grant. These sources of funding are gratefully acknowledged. Some parts of this chapter have appeared in Nichols (2015). Permission from Brill Academic Publishers to reprint those parts is gratefully acknowledged. The author owes many debts of many kinds for assistance in telling the story recorded in this chapter. See Nichols (2015) for a full list of thanks. Special thanks go to Kelly James Clark for insights and edits.

References

Barrett, K.C. (1995). A Functionalist Approach to Shame and Guilt. In *The Self-Conscious Emotions: Theory and Research*, edited by Jessica L. Tracy, Richard W. Robins, and June Price Tangney, pp. 25–63. New York: Guilford Press.

Bedford, O. and Hwang, K.-K. (2003). Guilt and Shame in Chinese Culture: A Cross-Cultural Framework from the Perspective of Morality and Identity. *Journal for the Theory of Social Behaviour* 33(2): 127–144.

Bellah, R.N. (2011). *Religion in Human Evolution: From the Paleolithic to the Axial Age*. Cambridge, MA: Belknap Press of Harvard University Press.

Betzig, L.L. (1982). Despotism and Differential Reproduction: A Cross-Cultural Correlation of Conflict Asymmetry, Hierarchy, and Degree of Polygyny. *Ethology and Sociobiology* 3(4): 209–221.

Betzig, L. (1995). Medieval Monogamy. *Journal of Family History* 20(2): 181–216.

Blumenthal, E. (1977). Models in Chinese Moral Education: Perspective from Children's Books. Dissertation Abstracts International. Retrieved from University Microfilms No. 77–7876.

Bond, M.H. (1986). *The Psychology of the Chinese People*. Hong Kong: Oxford University Press.

Boyd, R. and Richerson, P. (1988). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Chan, W. (1969). *A Source Book in Chinese Philosophy*. Princeton: Princeton University Press.

Chen, X. (2002). Social Control in China: Applications of the Labeling Theory and the Reintegrative Shaming Theory. *International Journal of Offender Therapy and Comparative Criminology* 46(1): 45–63.

Ch'ü, T.-T. (1961). *Law and Society in Traditional China*. Berlin: Mouton de Gruyter.

Clark, A. and Chalmers, D.J. (1998). The Extended Mind. *Analysis* 58: 7–19.

Cohen, D. and Gunz, A. (2002). As Seen by the Other…: Perspectives on the Self in the Memories and Emotional Perceptions of Easterners and Westerners. *Psychological Science* 13(1): 55–59.

Cole, P.M., Bruschi, C.J., and Tamang, B.L. (2002). Cultural Differences in Children's Emotional Reactions to Difficult Situations. *Child Development* 73(3): 983–996.

Cook, S. (2004). The Debate over Coercive Rulership and the "Human Way" in Light of Recently Excavated Warring States Texts. *Harvard Journal of Asiatic Studies* 64(2): 399–440.

Csikszentmihalyi, M. (2004). *Material Virtue: Ethics and the Body in Early China*. Leiden, Boston: Brill.

Eisenberger, N., Lieberman, M., and Williams, K. (2003). Does Rejection Hurt? An fMRI Study of Social Exclusion. *Science* 302: 290–292.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion* 6: 169–200.

Elison, J. (2005). Shame and Guilt: A Hundred Years of Apples and Oranges. *New Ideas in Psychology 23*(1): 5–32.

Fessler, D. (1999). Toward an Understanding of the Universality of Second Order Emotions. In *Beyond Nature or Nurture: Biocultural Approaches to the Emotions*, edited by A. Hinton, pp. 75–116. New York: Cambridge University Press.

Fessler, D. (2004). Shame in Two Cultures: Implications for Evolutionary Approaches. *Journal of Cognition and Culture* 4(2): 207–262.

Fessler, D. (2007). From Appeasement to Conformity: Evolutionary and Cultural Perspectives on Shame, Competition, and Cooperation. In *The Self-Conscious Emotions: Theory and Research*, edited by Jessica L. Tracy, Richard W. Robins, and June Price Tangney, pp. 174–193. New York: Guilford Press.

Fingarette, H. (1972). *Confucius: The Secular as Sacred*. Long Grove: Waveland Press.

Fournier, M.A., Moskowitz, D.S., and Zuroff, D.C. (2002). Social Rank Strategies in Hierarchical Relationships. *Journal of Personality and Social Psychology* 83(2): 425–433.

Frijda, N.H. (1993). The Place of Appraisal in Emotion. *Cognition and Emotion* 7(3–4): 357–387.

Fung, H. (1999). Becoming a Moral Child: The Socialization of Shame among Young Chinese Children. *Ethos* 27(2): 180–209.

Fung, H., Lieber, E., and Leung, P.W.L. (2003). Parental Beliefs on Shame and Moral Socialization in Taiwan, Hong Kong, and the United States. In *Progress in Asian Social Psychology: Conceptual and Empirical Contributions*, edited by Kuo-Shu Yang, Kwang-Kuo Hwang, Paul Pedersen, and Ikuo Daibo, pp. 83–109. Westport, Conn: Praeger.

Garrott, J.R. (1995). Chinese Cultural Values: New Angles, Added Insights. *International Journal of Intercultural Relations* 19(2): 211–225.

Gilbert, P. (1997). The Evolution of Social Attractiveness and its Role in Shame, Humiliation, Guilt and Therapy. *British Journal of Medical Psychology* 70(2): 113–147.

Gilbert, P. (1998). What is Shame? Some Core Issues and Controversies. In *Shame: Interpersonal Behaviour, Psychopathology, and Culture*, edited by Paul Gilbert and Bernice Andrews, pp. 3–38. Oxford: Oxford University Press.

Gilbert, P. (2003). Evolution, Social Roles, and the Differences in Shame and Guilt. *Social Research* 70(4): 1205–1230.

Gilbert, P. and McGuire, M. (1998). Shame, Social Roles and Status: The Psychobiological Continuum from Monkey to Human. In *Shame: Interpersonal Behaviour, Psychopathology, and Culture*, edited by Paul Gilbert and Bernice Andrews, pp. 99–125. Oxford: Oxford University Press.

Greer, S. and Lim, T.P. (1998). Confucianism: Natural Law Chinese Style? *Ratio Juris* 11(1): 80–89.

Griffiths, P. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Heine, S.J. (2001). Self as Cultural Product: An Examination of East Asian and North American Selves. *Journal of Personality* 69(6): 881–906.

Henrich, J. and Boyd, R. (1998). The Evolution of Conformist Transmission and the Emergence of Between-Group Differences. *Evolution and Human Behavior* 19(4): 215–241.

Henrich, J. and Gil-White, F.J. (2001). The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society* 22(3): 165–196.

Henrich, J., Heine, S.J., and Norenzayan, A. (2010). The Weirdest People in the World? *Behavioral and Brain Sciences* 33(2–3): 61–83.

Higham, C. (1988). Prehistoric Metallurgy in Southeast Asia: Some New Information from the Excavation of Ban Na Di. In *The Beginning of the Use of Metals and Alloys*, edited by Robert Maddin. Cambridge, MA: MIT Press.

Ho, P.-T. (1962/2008). *Ladder of Success in Imperial China: Aspects of Social Mobility, 1368–1911.* New York: ACLS Humanities E-Book.

Keltner, D. and Buswell, B. (1997). Embarrassment: Its Distinct Form and Appeasement Functions. *Psychological Bulletin* 122(3): 250–270.

Keltner, D. and Harker, L.A. (1998). The Forms and Functions of the Nonverbal Signal of Shame. In *Shame: Interpersonal Behaviour, Psychopathology, and Culture*, edited by Paul Gilbert and Bernice Andrews, pp. 78–98. New York: Oxford University Press.

Keltner, D., Haidt, J., and Shiota, M. (2006). Social Functionalism and the Evolution of the Emotions. In *Evolution and Social Psychology*, edited by Mark Schaller, Jeffry A. Simpson, and Douglas T. Kenrick, pp. 115–142. London: Psychology Press.

Kitayama, S., Markus, H.R., and Kurokawa, M. (2000). Culture, Emotion, and Well-Being: Good Feelings in Japan and the United States. *Cognition and Emotion* 14(1): 93–124.

Kitayama, S., Mesquita, B., and Karasawa, M. (2006). Cultural Affordances and Emotional Experience: Socially Engaging and Disengaging Emotions in Japan and the United States. *Journal of Personality and Social Psychology* 91(5): 890–903.

Kitayama, S, Markus, H.R., Matsumoto, H., and Norasakkunkit, V. (1997). Individual and Collective Processes in the Construction of the Self: Self-Enhancement in the United States and Self-Criticism in Japan. *Journal of Personality and Social Psychology* 72(6): 1245–1267.

Lewis, M.E. (1989). *Sanctioned Violence in Early China*. Albany: State University of New York Press.

Lewis, M.E. (2010). *The Early Chinese Empires: Qin and Han*. Cambridge, MA: Belknap Press of Harvard University Press.

Li, J., Wang, L., and Fischer, K. (2004). The Organisation of Chinese Shame Concepts. *Cognition and Emotion* 18(6): 767–797.

Lieber, E., Fung, H., and Leung, P.W.-L. (2006). Chinese Child-Rearing Beliefs: Key Dimensions and Contributions to the Development of Culture-Appropriate Assessment. *Asian Journal of Social Psychology* 9(2): 140–147.

Markus, H. and Kitayama, S. (1991). Culture and the Self: Implications for Cognition, Emotion, and Motivation. *Psychological Review* 98(2): 224–253.

Mauro, R., Sato, K., and Tucker, J. (1992). The Role of Appraisal in Human Emotions: A Cross-Cultural Study. *Journal of Personality and Social Psychology* 62(2): 301–317.

McLeod, K.C.D. and Yates, R.D.S. (1981). Forms of Ch'in Law: An Annotated Translation of The Feng-chen shih. *Harvard Journal of Asiatic Studies* 41(1): 111.

Menon, U. and Shweder, R.A. (1994). Kali's Tongue: Cultural Psychology, Cultural Consensus and the Meaning of "Shame" in Orissa, India. In *Emotion and Culture: Empirical Studies of Mutual Influence*, edited by Shinobu Kitayama and Hazel Rose Markus, pp. 241–284. Washington, DC: American Psychological Association.

Miller, P.J., Wiley, A.R., Fung, H., and Liang, C.-H. (1997). Personal Storytelling as a Medium of Socialization in Chinese and American Families. *Child Development* 68(3): 557–568.

Muller, A.C. (1991/2013). The Doctrine of the Mean (Zhongyong) 中庸. Available from: http://www.acmuller.net/con-dao/docofmean.html (last accessed July 15, 2015).

Nelson, L.J. (2006). Relations between Chinese Mothers' Parenting Practices and Social Withdrawal in Early Childhood. *International Journal of Behavioral Development* 30(3): 261–271.

Nichols, R. (2011). A Genealogy of Early Confucian Moral Psychology. *Philosophy East and West* 61(4): 609–629.

Nichols, R. (2013). The Origins and Effects of Filial Piety (Xiao 孝): How Culture Solves an Evolutionary Problem for Parents. *Journal of Cognition and Culture* 13(3–4): 201–230.

Nichols, R. (2015). Civilizing Humans with Shame: How Early Confucians Altered Inherited Evolutionary Norms through Cultural Programming to Increase Social Harmony. *Journal of Cognition and Culture*.

Parker, S.T. (1998). A Social Selection Model for the Evolution and Adaptive Significance of Self-Conscious Emotions. In *Self-Awareness: Its Nature and Development*, in Michel Ferrari and Robert J. Sternberg, pp. 108–134. New York: Guilford Press.

Pines, Y. (2002). *Foundations of Confucian Thought: Intellectual Life in the Chunqiu Period (722–453 B.C.E.)*. Honolulu: University of Hawai'i Press.

Ren, X. (1997). *Tradition of the Law and Law of the Tradition: Law, State, and Social Control in China*. Westport: Greenwood Press.

Rosemont, H. and Ames, R. (2009). *The Chinese Classic of Family Reverence: A Philosophical Translation of the Xiaojing*. Honolulu: University of Hawai'i Press.

Schurmann, F. (1968). *Ideology and Organization in Communist China*. Berkeley: University of California Press.

Shaver, P.R., Murdaya, U., and Fraley, R.C. (2001). Structure of the Indonesian Emotion Lexicon. *Asian Journal of Social Psychology* 4(3): 201–224.

Shaver, P., Wu, S., and Schwartz, J. (1992). Cross-Cultural Similarities and Differences in Emotion and its Representation: A Prototype Approach. *Review of Personality and Social Psychology* 13: 175–212.

Simoons, F. 1969. Primary Adult Lactose Intolerance and the Milking Habit: A Problem in Biological and Cultural Interrelations. I. Review of the Medical Research. *American Journal of Digestive Diseases* 14: 819–836.

Simoons, F. 1970. Primary Adult Lactose Intolerance and the Milking Habit: A Problem in Biological and Cultural Interrelations. II. A Culture Historical Hypothesis. *American Journal of Digestive Diseases* 15: 695–710.

Slingerland, E. (trans.) (2003). *Confucius' Analects: With Selections from Traditional Commentaries*. Indianapolis: Hackett.

Stipek, D. (1998). Differences between Americans and Chinese in the Circumstances Evoking Pride, Shame, and Guilt. *Journal of Cross-Cultural Psychology* 29(5): 616–629.

Summers, K. (2005). The Evolutionary Ecology of Despotism. *Evolution and Human Behavior* 26(1): 106–135.

Sznycer, D., Takemura, K., Delton, A. W., Sato, K., Robertson, T., Cosmides, L., and Tooby, J. (2012). Cross-Cultural Differences and Similarities in Proneness to Shame: An Adaptationist and Ecological Approach. *Evolutionary Psychology: An International Journal of Evolutionary Approaches to Psychology and Behavior* 10(2): 352–370.

Tooby, J. and Cosmides, L. (2008). The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables. In *Handbook of Emotions*, 3rd edn., edited by Michael Lewis, Jeannette M. Haviland-Jones, Lisa Feldman Barrett, pp. 114–137. New York: Guilford Press.

Triandis, H., McCusker, C., and Hui, H. (1900). Multimethod Probes of Individualism and Collectivism. *Journal of Personality and Social Psychology* 59(5): 1006–1020.

van der Sprenkel, S. (1966/2004). *Legal Institutions in Manchu China: A Sociological Analysis*. Oxford: Berg Publishers.

Van Norden, B. (trans.) (2008). *Mengzi: With Selections from Traditional Commentaries*. Indianapolis: Hackett.

Wong, Y. and Tsai, J. (2007). Cultural Models of Shame and Guilt. In *The Self-Conscious Emotions: Theory and Research*, edited by J.L. Tracy, R.W. Robins, and J.P. Tangney, pp. 209–223. New York: Guilford Press.

Zhang, X. (2004). A Restorative Justice Audit of the Chinese Criminal Justice System, PhD thesis. London: London School of Economics.

Zhong, J., Wang, A., Qian, M., Zhang, L., Gao, J., Yang, J., and Chen, P. (2008). Shame, Personality, and Social Anxiety Symptoms in Chinese and American Nonclinical Samples: A Cross-Cultural Study. *Depression and Anxiety* 25(5): 449–460.

Index

a priori/a posteriori, 3, 5, 20, 21, 35, 38, 75, 79, 86–89, 109, 125, 126, 140, 145, 154, 155, 159, 160, 162, 164–167, 169, 172, 180, 183, 184, 186, 200, 201, 220–225, 227–232, 235, 262, 354, 355, 362, 378, 383, 413, 490, 496, 503, 505, 507, 524
Adams, Robert, 210, 217–219
adṛṣṭa, 499
Advaita Vedanta non-naturalism, 496
Agent causation, 22, 63, 66, 308, 319, 322, 327–332
Alcoff, Linda, 44
Anthropology of Science, 101, 104
antinaturalism, 34–38, 43–45, 125, 126, 133, 134
antipsychologism, 35
anti-realism, 20, 76, 78, 79, 82–84, 152, 153, 267, 273, 274, 416, 419, 428, 432
anti-realism about truth, 268–270
Aristotle, 11–12, 23, 27, 94, 110, 168, 242, 444, 449, 497
arithmetical facts, 282, 284–287
Armstrong, David, 169, 176, 188
atomism, 497–500
authority, 3, 8, 9, 11, 12, 29, 46, 78, 80, 101, 106, 130–133, 171, 198, 199, 235, 279, 302, 338, 339, 343, 344, 374, 411, 418, 419, 421, 468, 471, 506, 507, 514, 520
Barkow, J., 71
Barrett, Justin, 99, 107, 466, 467, 469–471, 485, 486, 490
belief, justified, 50, 83, 175, 258, 293, 398, 421, 430, 440, 482, 484, 491, 506, 507
beliefs, 2, 4, 5, 7–9, 13, 18, 21, 22, 24, 26, 29, 30, 34, 40, 51, 61, 65, 66, 71, 72, 79, 83, 87, 89, 94, 105, 115–117, 120, 127, 129, 131, 134, 154, 163, 164, 173, 177, 197, 205, 210, 223, 227, 229, 235, 247–255, 258–260, 264, 265, 271, 301, 302, 309, 324, 348, 364, 371, 372, 375–379, 383, 386–399, 412, 418, 431, 432, 437–440, 443, 444, 448, 456, 462–465, 471, 472, 474, 475, 481–491, 503, 506, 507, 522
Bergson, Henri, 35, 39
Bering, Jesse, 472
Berkeley, Bishop George, 209, 214
Blackburn, Simon, 9, 10, 97, 337, 342, 355–357
Boyer, Pascal, 465–471, 476, 485, 490
brains-in-a-vat argument (BIVA), 154–155, 159
Brosnan, Kevin, 372, 377
Butler, Judith, 44, 45
by-product theory, 466–471

Canberra Plan, 230, 379–381
Carruthers, Peter, 68–71
catastrophism, 111
categorical imperative, 21, 29, 344, 418

The Blackwell Companion to Naturalism, First Edition. Edited by Kelly James Clark.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

causation
  agent causation, 22, 63, 66, 308, 319, 322, 327–332
  event causation, 308, 309, 327–329
  substance causation, 88, 322, 327–331
causation, theories of
  Nyāya, 495–496
  Sāmkhya, 495, 497, 501, 503
Chalmers, David, 3, 5, 161, 163, 164, 192, 215, 263, 515
character, 2, 10, 12, 16, 23, 24, 27–29, 43, 51, 59, 62, 92, 93, 96, 98–100, 103, 105, 120, 126, 128, 136, 137, 139, 140, 151–153, 160, 163, 168, 169, 172, 175, 178, 183, 193, 196, 197, 202, 204, 206, 209–212, 214, 218, 221, 224, 226, 230, 231, 240, 241, 247, 248, 251, 252, 259, 272, 273, 277, 281, 292, 310, 328, 338, 345, 346, 365, 370, 375, 389, 401, 404, 406, 408, 409, 414, 417, 424, 425, 428–433, 438–440, 445, 465, 466, 469, 474, 476, 482, 494–496, 501, 502, 506, 508, 516, 518, 522, 467468
charismatic ecologies, 473, 474
Christianity, 110, 113, 436, 449, 483, 484, 488
Churchland, Paul, 4, 5, 204, 205, 207, 211, 387, 433
Clarke, Randolph, 308–309, 321
Clark, K.J., 1–14, 148, 179
closure, 172, 188, 189, 191–194, 201, 308, 309, 315, 316, 322, 323, 500
cognitive labor, 130, 132–133
cognitive science, 26, 30, 37, 64, 70, 200, 223, 250, 312, 326, 463–466, 470, 471, 475, 481–482, 485–486
Cohen, Emma, 471
coherence, 61, 134, 153, 196, 252, 253, 256, 287, 321, 428, 507
Colyvan, Mark, 222, 293, 295–297
compatibilism, 306, 307, 311, 313, 314, 321, 324–326, 332, 509
concepts, 3, 17, 35, 77, 92, 110, 124, 140, 151, 171, 183, 196, 211, 221, 237, 246, 271, 281, 301, 316, 324, 342, 351, 369, 387, 403, 419, 436, 447, 464, 482, 497, 513
conceptual analysis, 20, 77, 88, 124, 151, 171, 230, 232, 351
consciousness, 4, 5, 7, 8, 13, 31, 36, 37, 160–162, 164, 166, 175, 190, 192, 199, 200, 206, 210, 211, 213, 219, 236–240, 242, 263, 310, 316, 326, 445, 446, 468, 495–497, 500, 502, 503
constitutively explain, 406–409, 412
constructive empiricism, 67, 104, 105
constructive interpretation, 100–107
convention T, 265
Copp, David, 10, 353, 402, 413
Cornell Realism, 362
correspondence theory of truth, 248–251
Cosmides, L., 71, 467, 513
Coyne, Jerry, 323–326
creation, 41, 79, 93, 110–112, 117, 118, 120, 142, 178, 200, 201, 206, 214, 295, 297, 458, 486, 498, 502
creationism, 117, 118, 120, 486
credibility-enhancing displays (CREDs), 474–475
cultural group selection (CGS), 474–475
Cuneo, Terence, 10, 401–415

Darwin, Charles, 6–7, 13, 19, 23, 24, 39, 43, 80, 109–114, 120, 177, 178, 200, 201, 206, 243, 244, 370, 377, 387, 389, 390, 463
Davidson, Donald, 188, 189, 251, 271, 358
Dawkins, Richard, 6, 9, 63, 114, 116, 177, 178, 435, 436, 440, 444, 445, 462
debunking arguments, 372, 373, 375–379, 383, 395–399
Dedekind–Peano axioms, 283–284
deflationary (theory of truth), 28, 39, 77, 86, 161, 246, 253–257, 263, 270–273
deflationary theory, 256, 270–272
Deleuze, Gilles, 35–38, 40–43
Dennett, Daniel, 13, 70, 178, 319, 325, 371, 435, 443–445, 448, 449, 462, 474
deontology, 345, 346, 417–419, 431
Descartes, Rene, 6, 18, 22, 93, 153–155, 184, 185, 198, 199, 202, 218, 224, 236, 240, 324
determinism, 22, 41, 133, 192, 306–311, 313–316, 319–326, 331–332
Dewey, John, 7, 23, 30, 234–244
Doris, John, 416, 428–432
Dougherty, Trent, 67
Dretske, Fred, 210, 213
dual inheritance, 474
dualism, mind-body, 2, 20, 489–491
dual process theory, 346, 416–419
Dummett, Michael, 152, 153, 257, 264

duty, 30, 41, 104, 242, 346–348, 382, 440, 454, 517
Dworkin, Ronald, 102, 105, 454

Ebbinghaus, Herman, 199
Eccles, John, 192
education, 30, 58, 117, 118, 120, 161, 199, 423, 435, 487
eliminative materialism, 203–207
emotion, 9, 11, 12, 26, 72, 73, 199, 202, 206, 216, 336–340, 344–347, 371, 380, 387, 389, 390, 397, 417, 418, 420–425, 437, 438, 456, 467, 472, 473, 483, 485, 512–523
empirical science, 51, 58, 59, 86, 124, 173, 292, 295–298, 494, 505
Enoch, David, 355, 365, 372, 377, 401
epistemic (theory of truth), 254
epistemically pure, 130
epistemically sullied, 130
epistemic normativity, 51, 55, 56, 220, 224, 228
Epistemic Police, 101–103
epistemology, 7, 18–20, 24–27, 50–52, 94, 95, 104, 124, 125, 129–132, 135, 166, 169, 175, 198, 220, 222–225, 227, 228, 230–232, 234–238, 244, 252, 295, 298, 299, 351, 355, 481, 504–508
epistemology, modest, 135
error theory, 9, 10, 357, 376, 378, 428–432
evolution and ethics, 10–13
evolutionary argument against naturalism, 64, 115–117, 140
evolutionary debunking arguments, 372, 373, 378
evolutionary psychology, 25, 54–55, 71, 395, 443, 467, 471
evolutionary theory, 63–65, 110, 113–116, 178, 179, 482
expressivism, 9, 10, 14, 356–358, 366, 413
extinction, 111

Feigl, Herbert, 188, 202
feminism, 43, 49–59
Ferguson, Andrew, 447, 448
Fischer, John Martin, 307, 311–312, 321, 332
FitzPatrick, William, 386–399, 409, 410, 413, 414
folk language, use of by naturalists, 448, 449, 461
Fraser, Benjamin, 372, 379–382
free will, 5, 13, 14, 22, 23, 40, 173, 174, 185, 199, 206, 305–316, 319–332

Gibbard, Allan, 9, 337, 342, 356, 358, 401, 403
God, 2–6, 8, 9, 13, 17–19, 21, 22, 27, 37, 41, 43, 61–63, 93, 109–117, 120, 121, 124, 125, 127, 136, 139, 140, 142–147, 169, 176–180, 183, 191, 200, 201, 209, 214, 215, 217–219, 235, 239, 241, 268, 278, 279, 281, 302, 319, 322, 352, 369, 395, 403, 406, 416, 437, 442, 443, 449, 457–461, 467, 468, 470–473, 475, 476, 482–485, 487–491, 496, 497, 499, 500, 507, 509, 510
Gopnik, A., 71
Gould, Stephen Jay, 43, 111, 117, 199
Greene, Joshua, 12, 19, 345–347, 387, 416–419
Guattari, Félix, 42, 43
Guthrie, Stewart, 466, 485, 486

Hall, G. Stanley, 199
Harman, Gilbert, 364, 365, 375, 376, 378, 387, 389, 428–432
Harris, Sam, 324–326
von Helmholtz, Hermann, 199
Hempel, Carl, 81, 137, 143, 187–188
Henrich, Joseph, 474, 475, 514–516, 520
Heyman, Gene, 313
Hobbes, Thomas, 6, 41, 124, 352
Hodgson, David, 308, 310, 314
Horwich, Paul, 254–257, 270
human nature, attempts to ground natural human rights on, 11–12
Hume, David, 6, 16, 22, 25–27, 39, 40, 42, 50, 51, 55, 58, 77, 84, 336, 343, 390, 435, 436, 462, 481, 483, 484
Hursthouse, Rosalind, 11, 12
Husserl, Edmund, 34–37, 45, 125
hypothesis of the bodies of common sense, 99, 100

ideology critique, 52, 53, 56–59
image of god, attempts to ground human rights on, 458, 459
incompatibilism, 312, 314, 315, 320–322, 326, 332
indispensability argument, 292–297
instrumentalism, 82, 95–100, 105, 228
integrative naturalism, 91–95, 100–101
intelligent design theory, 118–120, 143, 200, 201, 206
intentionality, 210, 212–214
interpretation, 6, 39, 50, 76, 80, 84–89, 96, 100–107, 117, 120, 121, 127, 139, 156,

158, 159, 168, 209, 210, 215, 231, 242, 252, 309, 315, 378, 418, 430, 437, 439, 441, 445, 464, 487
intuitions, 3, 35, 65, 82–84, 89, 94, 139, 151, 152, 155, 159–166, 170–172, 175, 176, 180, 190, 226, 227, 229, 231, 235, 239, 269, 272, 291, 312, 313, 332, 340, 345–347, 356, 359, 364, 373, 418, 419, 453–457, 467
intuitive ontologies, 468, 469, 486

Jackson, Frank, 189, 192, 193, 230, 255, 272, 355, 356, 358, 360–362, 365, 379, 401, 405, 406, 409, 412
James, William, 168, 169, 199
Johnson, Phillip E., 200, 201
Jonas, Hans, 35, 37–38, 126
Joyce, Richard, 9, 11–13, 16, 369–383, 387, 390, 395, 397, 399, 419, 427, 431

Kahneman, Daniel, 116, 162, 163, 313
Kane, Robert, 309, 310, 314, 320, 331
Kelemen, Deborah, 472, 486
Kitcher, Philip, 8, 11, 17, 18, 80, 84, 124, 125, 129–133, 178, 179, 234, 244, 371, 387, 390, 435–446, 505
knowledge, 2, 7, 14, 17, 28–31, 36, 40, 41, 43, 62, 64, 76, 77, 79, 82, 85, 89, 93, 99, 100, 120, 125, 127, 133–135, 159, 171, 197, 202, 220–223, 225–227, 229, 230, 234–244, 289, 291, 300, 301, 518
knowledge production, 120, 125, 133–135, 243
Kornblith, Hilary, 25, 51, 225, 226, 503
Korsgaard, Christine, 341, 344, 352, 401
Kripke, Saul, 164, 183, 266, 355, 363

laboratory life, 126–129
Latour, Bruno, 79, 125–129, 131–134
laws of nature, 112, 114, 120, 142, 143, 217, 219, 262, 306, 307, 320, 326
Lawson, E. Thomas, 464–467
Leibniz, Gottfried, 41, 184–186, 191
Levy, Neil, 305–316, 320, 322, 326, 332
Lewis, David, 158, 159, 174, 188, 189, 247, 248, 267, 279, 282, 353, 379, 517
libertarianism, 306–310, 313, 320–324, 326, 332
lightweight epistemological naturalism, 220, 227–232
literary inscriptions, 128
logical behaviorism, 202, 203
Longino, Helen, 50, 125, 126, 129, 133–135
Lowe, E.J., 172, 176, 192, 322, 328–330
Lynch, Michael, 246–260, 264, 273, 274

Mackie, J.L., 9, 16, 20, 25, 354, 357, 380, 381, 431
Maddy, Penelope, 78, 93, 95, 101, 107, 224, 225, 229, 278, 292, 294, 298–303
Mādhyamika criticism, 508
manifest image of world, 86, 204, 205, 448
marks of the natural, 411–413
Marx, Karl, 234, 239, 447
mathematical knowledge, 289, 290, 295, 297, 300
mathematical model, 137, 143, 144, 146, 147, 524
McCauley, Robert N., 462–476, 482, 486, 487
Melzoff, A., 71
Mendel, Johann Gregor, 111
Merleau-Ponty, Maurice, 35, 40, 44, 45
metaethics, 10, 14, 336–344, 351–366, 370, 372, 376, 380, 383, 395, 397–399, 416, 417, 420, 423, 425, 426, 428, 430, 431
metaphysical construction, 212
metaphysics, 3, 14, 38, 39, 41, 61–73, 76, 77, 85–89, 114, 146, 168–180, 203, 226, 246, 247, 258, 299–301, 315, 324, 327–329, 362, 396, 422, 496, 505, 510
  analytic, 169–176
  naturalistic, 61–73, 77, 91, 142, 153, 319
  nonnaturalistic, 168–180
  theistic, 142, 176–180
Mīmāmsā ritualism, 509
minimally counterintuitive (MCI) representations, 469, 470
model-theoretic argument (MTA), 153, 155–159
mokṣa, 495, 508, 509
moral dumbfounding, 340
moral explanations, 364, 365
moral intuitions, 345, 346
moral judgment/consequentialist, deontological, 10, 11, 336–346, 364–375, 378, 379, 383, 387–390, 417–423, 425
moral knowledge, 17, 30, 379, 390, 395, 397
moral nativism, 371–383
moral naturalism, 369–383, 395, 401, 402, 405, 499, 500, 508–510
moral nonnaturalism, 414
moral realism, 9, 10, 373, 374, 378, 380, 388, 394, 395, 397–399, 418, 422
moral reasoning, 206, 341, 417

moral relativism, 419–428
moral skepticism, 16, 397
mythology, 127, 238, 241, 244

Nagel, Thomas, 64, 179, 189, 211, 231, 448
natural human rights
  idea of explained, 369
  idea of grounding explained, 17, 21, 29, 280, 287, 451–453, 457, 460, 461
  secular attempts at grounding, 446
  theistic grounding, 457, 460
naturalism, 1, 16, 34, 49, 61, 75, 91, 109, 124, 136, 150, 168, 182, 196, 209, 220, 234, 246, 262, 277, 289, 305, 319, 336, 351, 369, 386, 401, 416, 435, 447, 462, 481, 494, 512
naturalism (metaphysical or ontological), 291
naturalism about truth, 262–275
naturalism, affirmative, 34, 35, 38–43
naturalism, methodological, 1, 5, 6, 34, 109, 113–116, 119, 120, 136–147, 170–175, 177, 179, 180, 201, 203, 221–223, 225, 227, 228, 262, 305, 354, 369, 370, 381, 496, 503–505, 509
naturalistic explanations of morality, 13, 56, 57, 120, 177, 200, 376, 483
naturalized epistemology, 18, 25, 26, 52, 505
natural selection, 11, 23, 109, 112–117, 120, 121, 177–179, 377, 387, 389, 395–397, 463, 472, 474, 490, 514
nature, 2–7, 11, 12, 20, 25, 29, 40–43, 45, 77, 95–100, 388–390, 517
negative experimental philosophy, 220, 227, 232
neo-Kantianism, 35
network analysis, 356, 360–362
Newtonian mechanics, 82, 98–100, 185
Newton, Isaac, 19, 95, 98–100, 184–186, 198, 414
niche construction, 31, 472–473
Nichols, Shaun, 314, 325, 338, 339, 348, 416, 419–425
Nietzsche, Friedrich, 35, 38–40, 42, 390
nonarithmetical facts, 282, 284
noncognitivism, 371, 390, 416–419, 425, 428
Norenzayan, Ara, 470, 472, 473, 475, 520
normative conception of truth, 135, 354, 361–362l
normativity, 10, 24–25, 28, 50–59, 220, 224, 228, 246, 247, 258, 259, 380
Nyāya pramānavāda (epistemology), 504–508

ontological naturalism
  hard
    the Ājīvika-s, 497
    the Cārvāka-s, 494, 496, 497, 500
    the svabhāva-vādins, 496, 497, 500, 501
  mitigated (Nyāya-Vaiśeṣika), 496
  protonaturalism (Sāmkhya), 501–503
  qualified (Bauddha), 494
open question argument, 28, 351–354, 356, 359, 360, 363, 365
Oppenheim, Paul, 137, 188

Papineau, David, 182–194, 196, 202, 205, 221, 229, 250, 262, 315, 322, 503, 504
Parfit, Derek, 10, 365, 366, 380, 401, 402
parsimony, 354, 374, 376, 396–399, 490
particularism, 413
Pavlov, Ivan, 199, 203
Peirce, Charles Sanders, 251, 252, 264
Pennock, Robert T., 119, 120, 138, 200, 201, 203
personhood, attempts ground natural human rights on, 456
phenomenological psycholog, 37
philosopher-monarch, 129, 130, 132, 133
philosophy of science, 50, 66, 67, 75–89, 100–107, 125, 143, 209, 329, 351, 362–364
physics, 3, 19, 22, 35–39, 41, 66, 87, 88, 91, 99, 109, 113, 134, 137, 140, 146, 171–173, 182, 185, 187–189, 193, 198, 204, 205, 211, 214, 215, 218, 243, 247, 270, 278, 294, 298, 299, 306–313, 315, 322–324, 396, 403, 404, 412, 414, 443, 448, 469, 497, 503, 504
Plantinga, Alvin, 18, 64, 73, 115–117, 140, 145, 176–177, 179, 222, 268–270, 406, 490
pluralism, 30–31, 88, 125, 133, 135, 185, 257–258, 273–275, 494
pluralist (theory of truth), 246, 257, 273–274
Popper, Karl, 101, 125, 192
positivism, logical, 2, 118
practical reason, 21, 22, 312, 344, 423
Prinz, Jesse, 28, 337, 339, 416, 423–428
provisional atheism, 139140
psychological behaviorism, 203
psychopathy, 338, 339
Puruṣārtha-s, 508
Putnam, Hilary, 20, 24, 25, 81, 96, 151–159, 188, 251, 274, 277, 292–297, 355, 363

Quine, W. V. O., 7, 25, 50, 51, 58, 92, 94, 95, 159, 170, 222–225, 229, 234, 271, 272, 277, 290, 292–296, 298, 363, 364, 505, 508

Railton, Peter, 13, 29, 354, 412
rational agency, attempts to ground natural human rights on, 327, 454–457
rationalism (about moral judgment), 336, 338, 339
realism, moral, metaphysical, naturalistic, scientific, 24–27, 61–73, 75–89, 91–107, 150–166, 168–180, 277–287, 336–348, 369–383, 401–433
Rea, Michael, 6–7, 64, 65, 139, 170, 174, 177, 268
reasons, 4, 17, 37, 51, 61, 76, 93, 113, 124, 137, 151, 169, 185, 197, 211, 222, 248, 270, 278, 290, 308, 320, 336, 352, 370, 387, 403, 417, 436, 451, 482
reductionism, 329, 376, 401
reference (theory of), 159, 173, 249, 250, 275
reliability, 64–68, 70–73, 116, 129, 159, 160, 162, 175, 177, 252, 299, 301, 396, 506, 508
religion, 7, 14, 17, 39, 112, 124, 141, 169, 177, 180, 199, 241, 481–482, 485–486
representation, 79, 80, 134, 140, 144, 153, 154, 159, 211, 212, 214, 236–239, 248–250, 257, 260, 312, 371, 374, 377, 380, 398, 425, 456, 463–470
Richards, Robert, 12, 38, 39
rights, human, 13, 447–461
Roland, Jeffrey, 289–303
Rorty's pragmatism, 271
Rosenberg, Alex, 2, 17, 91, 292
Rta, 509
Ruse, Michael, 116, 138, 373, 374, 378
Russell, Bertrand, 146, 197–199, 248, 253, 265, 267, 435, 436
Ryle, Gilbert, 5, 202, 203

Schroeder, Mark, 343, 362, 406
Schroedinger, Erwin, 4
science, 3, 20, 36, 51, 64, 75, 92, 109, 124, 136, 150, 169, 182, 196, 211, 235, 247, 262, 279, 290, 305, 322, 354, 388, 403, 443, 448, 463, 481, 495
science, sociology of, 79, 124–135
science *vs.* metaphysics, 3, 14, 76–77, 85–89, 170–173, 178
science *vs.* religion, 112, 124, 141, 177, 199, 481–482, 485–486
scientific antirealism, 78, 79
scientific explanation of moral belief, 386–399
scientific naturalism about morality, 386–399
scientific realism, 66, 67, 76–85, 89, 92, 95, 97, 98, 100, 104, 105, 153
scientism, 34, 35, 91–107
second philosophy, 78, 292, 298–303
Sellars, Roy Wood, 1
Sellars, Wilfrid, 3, 86, 204
semantic primitivism, 275
sentimentalism, 336–338, 423–428
set theory, 158, 281, 282, 296–302
Shafer Landau, Russ, 9, 353–355, 365, 397, 401–403, 412, 414
signaling theory, 472–473, 475
situationism, 389
skepticism, 16, 44, 50, 78, 83, 84, 120, 153, 154, 175, 228, 238, 294, 320, 344, 397, 508
Smart, J.J.C., 188, 189
Sober, Elliott, 64, 118, 138, 377, 474
social construction, 45, 125, 126, 129
social epistemology, 129–132
society, 41, 57, 58, 68, 79, 120, 121, 125, 130, 242, 359, 387, 426, 427, 443–447, 457, 464, 465, 467, 470, 473, 475, 484, 485, 508, 514, 516–519, 522, 523
Spinoza, Benedict, 38–42
standards, 9, 12, 25, 27, 118, 234, 251, 277–278, 284, 322, 353, 360, 380, 410–415, 417, 422, 428, 450, 458, 471, 442447
Stein, Howard, 96–97
Sterelny, Kim, 379–382
Steward, Helen, 314–315, 321, 327, 328
Street, Sharon, 11, 13, 373–375, 378, 395, 397
strong ontological naturalism, 263, 265, 271
strong program, 29, 125, 126, 128, 133
structuralism, 284
Sturgeon, Nicholas, 10, 362, 364, 365, 375, 412
superassertibility, 252, 257
supernatural, 2, 3, 5, 6, 8, 16–20, 36, 38, 40, 49, 109–111, 113, 114, 118, 120, 121, 136, 139–142, 145, 168, 169, 200, 201, 221, 263, 279, 285, 287, 305, 322, 324, 331, 352, 440, 468, 481–483, 486
systematic, 66, 73, 76, 77, 84, 92, 93, 96–98, 138, 147, 172, 194, 198, 327, 353, 354, 364, 429, 467, 496

teleology, 39, 210, 211, 214, 486, 503
theism, 2, 19, 136, 139, 140, 142, 143, 176, 177, 209–210, 213–219, 441, 448, 457, 459, 484

theological incorrectness, 471
theory of mind, 468, 476, 488, 519
Tierney, Brian, 447
Timmons, Mark, 366, 380, 418
Tooby, J, 71, 467, 513
Tremain, Shelley, 44–45
trolley cases/footbride, footbridge switch, 348, 417
truth, 4, 10, 68, 72, 78, 82, 84, 113, 116, 117, 134, 137, 138, 147, 158, 165, 241, 243, 246–260, 262–274
Tversky, Amos, 116, 163, 313

underdetermination, 81, 82
Universal Declaration of Human Rights, 447, 450
universalization, 341
U.S. Declaration of Independence, 447

value (of true belief), 259
van Fraassen, Bas, 67, 77, 80–86, 91, 104, 105, 137, 139, 143
van Inwagen, Peter, 62, 114, 174, 178, 307, 314, 315, 320
Vidyā
  adhyātma, 495
  aparā, 495
  ānvīksikī, 495, 505
  parā, 495
virtue, 11, 20, 21, 23, 24, 26–30, 86, 102, 103, 105, 173, 210–212, 249, 250, 257, 258, 267, 272, 279–282, 284, 296, 299, 303, 314–316, 322, 328, 347, 370, 372, 376, 379, 381, 389, 396–398, 406, 408–410, 413, 425, 429–432, 437, 450, 453, 456, 459, 484, 495, 500, 501, 506, 516, 518

weak ontological naturalism, 262–265, 305, 322
Whitehouse, Harvey, 466, 467, 470, 474
Wielenberg, Erik, 372, 377, 378
Williamson, Timothy, 13, 160, 228, 251
Wilson, E.O., 10, 463, 469, 474
Wright, Crispin, 251, 252, 257, 268–271, 273, 274
Wundt, Wilhelm, 199

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.